

RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multichannel sequencer. *Genome Res.* 10 (11), 1757-1771 (2000)

Kono, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y. and Hayashizaki, Y.

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. *Genome Res.* 11 (2), 281-289 (2001)

Kondo, S., Shinagawa, A., Saito, T., Kiyosawa, H., Yamamaki, T., Aizawa, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and Hayashizaki, Y.

Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences. *Mamm. Genome.* 12, 673-677 (2001)

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

e mouse tissues.

FEATURES

Location/Qualifiers	Source
1. .628	

/organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="E230022602"
 /clone_1lib="RIKEN full-length enriched, 2 days pregnant
 adult female oviduct"
 /sex="female"
 /tissue_type="oviduct"
 /dev_stage="2 days pregnant adult"
 /lab_host="DH10B"
 /note="Site_1: SalI; site_2: BamHI; cDNA library was
 prepared and sequenced in Mouse Genome Encyclopedia
 Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in
 RIKEN. Division of Experimental Animal Research in Riken
 contributed to prepare mouse tissues. 1st strand cDNA was
 primed with a primer [5']
 GAGGAGAGAGAGCGCCGACCACTCGAGTTTCTTTTCTTTCVN 3']. cDNA was
 prepared by using trehalose thermo-activated reverse
 transcriptase and subsequently enriched for full-length by
 cap-trapper. Second strand cDNA was prepared with the
 primer adapter of sequence [5']
 GAGGAGAGAGATTCGCGTTAAATTAATCCGCCGCCGCC 3']. cDNA
 was cleaved with BamHI and XhoI. Vector: a modified
 bluescript KS(+) after bulk excision from Lambda PUC I."
 142 c 106 g 183 t 2 others

BASE COUNT	195 a	142 c	106 g	183 t	2 others
ORIGIN					

Query Match	12.7%;	Score 455.4;	DB 9;	Length 628;
Best Local Similarity	89.4%;	Pred. No. 3.4e-79;		
Matches 489; Conservative	0;	Mismatches 58;	Indels 0;	Gaps 0;

QY	555	gtcctcaatcatcaagaataaccttcgcgaagccgggagtlltttgaatatcttcacag	614
Db	82	GTACCTCAAGTCATTAACGAAATAACCTTCTCTPAAACCCAGGGAGTTTTTTGAATCATCTCCACAG	141
QY	615	actgaatgatgcgtgtaatttgaaagtaaacaaccgaagcgaattccccacaacattta	674
Db	142	ACTGGAATGGCTGATTAATTGAAGCAATCACCTCAGTCGAAATTTCCCACATCATTTTA	201
QY	675	tggactaatctctcttatctctcttagtacctgatgatnaacgtccaccccgttlactcta	734
Db	202	CGGACTAAATCTCTTATTCTTAGTGCGATGAATAATATGCGCCTCACCCGTTTGCCTGA	261
QY	735	taaacctctctgtcaacacatgcgaagaactacatttggccttgagaccttgaagcaaccatal	794
Db	262	CAAACCCTCTGTGACAGCAATGCCCGCCGCACTGGCTGGACTTTGAAGCAACCGTAT	321
QY	795	ccataatttaagaaatttgaccttattcttcctgcagtaattaactgyltttaglbgag	854
Db	322	CCATAATTTAAGAATTTGACTTTTATTCTCTGCAATAATTTAACTGTTGGTAATGAC	381
QY	855	gaaaacaaaattatcatcaattaaatgaaaaactcttttgcacccctccagaanaactgatga	914
Db	382	GAAACAACAAAATCAATTAATTAATGAACATGCAATTTCACACCTCCAGAACTAGATGA	441

QY 915 attgatttagaagaataaagattgaaatcttcacagcgtcattatcaaggacttaa 974

Db 442 ATTGGATTATTTAGGAATATTAAGATTGAAATTTCTCCACCGAACAATATTTAAGATCTGAA 501

QY 975 ggaagctgcacacatggaatctctcctataatccaatccagaagaattcaagcaaaccaat 1034

Db 502 GGAGCTGTGCGCAATTGAAATATTTCTTAATACCCAAATCCAGAAATTTGAAGTAAATCAATT 561

QY 1035 tgattatcctgtcaactcaagtcctcctaagcctagaaggaattgaattccaatatcca 1094

Db 562 CGATTGTCGTGGCAACCAACGACTCTCAGCGCTTGAAGGAANTTAATTAATATCCA 621

QY 1095 acaaaag 1101

Db 622 ACAAAAG 628

RESULT	2
AI089319/c	
LOCUS	AI089319 469 bp mRNA linear EST 18-AUG-1998
DEFINITION	qb16dd05.xl Soares-pregnant-uterus-NBHPU Homo sapiens cDNA clone
IMAGE:	1956425 3', mRNA sequence.
ACCESSION	AI089319
VERSION	AI089319.1 GI:3428378
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 469)	NCI-CCAP	http://www.ncbi.nlm.nih.gov/ncicgap		
	National Cancer Institute, Cancer Genome Anatomy Project (CGAP)			
	Tumor Gene Index			
	Unpublished (1997)			
	Contact: Robert Strausberg, Ph.D.			

FEATURES

Email: cgabbs-r@mail.nih.gov
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 444.
Location/Qualifiers

FEATURES
source

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1 IMAGE:1696425"
/clone_1db="Soares_pregnant_uterus_NBHPU"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: uterus; Vector: pT73-Pac; Site_1: Not I;
Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dN) primer [5',
AAGTGAAGAAATTCGCGCCCTTTTCTTTTCTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization. Library
constructed by M. Fatima Bonaldo."

```

BASE COUNT	156 a	78 c	70 g	165 t
ORIGIN				

Query Match	12.1%	Score 433.8;	DB 9;	Length 469;
Best Local Similarity	96.6%	Pred. No. 5.8e-75;		
Matches 435; Conservative	0;	Mismatches 6;	Indels 0;	Gaps 0;

Oy 311 caaacgctgctgtttttaagaacgacctaaagtgtttaaattcaccacttaagatggtg 3200
Db 469 CAAACGCTGCTGTTTAAAGACAGACCTAAGTGTGTTAATCACCACCTTGTGATGGTG 410
Oy 3201 aacttatggtgtgtgaatatctcgttaaagcagttaaaggaanaagctgtgaatgc 3260

Db 409 AATGTATGGTGTGTGAATATCTCAGTAAAGCAGTTAAAGAAAAAGAGCTGGAATGC 350
Qy 3261 acgatccaggaacttaattcaggaaggaaggctgtatgtacacattcacttaag 3320
|||||
Db 349 ACTGATTCAGCACTTAAATTCAGGAAGAAAGGTGTGTATGTACACATTTTCACCTTTAG 290
Qy 3321 cagaanaactctctcagaagaatgaacttaactctctctcttcgaagcgaagcagtgagat 3380
|||||
Db 289 CAGAAAAATCTTTCTTCAAGAAATGACTTTACTTCTCTTTCACATGCCAGACGATGAGAT 230
Qy 3381 actaactcttactagctgtctctctcagtcctcagtcctcagtcctcagtcctcagtcctc 3440
|||||
Db 229 ACTAACTTTTAACTACTGTCTTCTCTAGTCTCTAGTATATAGAAATTTTGCTTTC 170
Qy 3441 ataatggaacaccttaagcaggaagaaatgttttcagatagtttcaaatcnccaa 3500
|||||
Db 169 AATAATGGAACCTTTAAGCAGGAGAAAGAAATGTTTCAATAGTTTCAATACACCAA 110
Qy 3501 aaatgttcgacacaaataactcgaatcnaaccataatgcccttatgataatagat 3560
|||||
Db 109 AAATGTTGAAACACAAATACTGGAATCAACCATRATGCACTTATTTGAATATATAGT 50
Qy 3561 tgatagattgtctcgaaga 3581
|||||
Db 49 TGTATGATTTGTTCTGAAAA 29

RESULT 3 444 bp mRNA linear EST 13-DEC-1999
AI805323/c
LOCUS
DEFINITION
IMAGE:2087673 3', mRNA sequence.
AI805323
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS
TITLE
NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
JOURNAL
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 612 Std Error: 0.00
Seq primer: -400P from Gibco
High quality sequence stop: 437.
Location/Qualifiers

FEATURES
Source
1. .444
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2087673"
/clone_1b="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pRT3D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not 1; Site_2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NbHL19W, testis NHT, and B-cell
NCI-GCAP GC81) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo."

BASE COUNT 140 a 79 c 60 g 165 t

Query Match 11.7%; Score 420.4; DB 9; Length 444;
Best Local Similarity 99.3%; Pred. No. 2.5e-72;
Matches 443; Conservative 0; Mismatches 1; Indels 2; Gaps 2;

Qy 2319 gagaaactgtgggggtgctcagtagaggaattactgtgatgaagaatgatacacaata 2378
|||||
Db 444 GAGAAATCTGTGGGGGTCCTCATGAGGAGATTCTGTGTATG-AAATGAATACCAAAAA 386
Qy 2379 ttaattataataatagcctaagaataatatttcaagagacatgaggaataaataatg 2438
|||||
Db 385 TTAATTTAATAATATAGCTAAGATAAATATTTTCAAGGACATGAGGAAAAATGAAATG 326
Qy 2439 actaatgctcttacaagaaggagaatataataatgatatatattatgtgacatt 2478
|||||
Db 325 ACTATAGCTCTTACAAAGGGAAGTAATATATATATATATATATATATATATATATATAT 266
Qy 2499 ttgcataagaataatgaagaatactactcagtaacattcattcttctcaacatgca 2558
|||||
Db 265 TTG-ATAAGAAATTAAGAAATCTACTCTGTAACATTCATTTTCTAATGCA 207
Qy 2559 ttatgagtaaccactactatgtgcataagcattgcaatataagtcctggaagtagcaagt 2618
|||||
Db 206 TTTATTTGATACCCACTACTATGTGCATAGCATATATATATATATATATATATATAT 147
Qy 2619 gcaaaaccttcaatctgtagatagtggttaatagaacaaagacataaagctcactg 2678
|||||
Db 146 GCAGAACCTTTCATATCTGTATAGTGTTTAATATACAAAGACATATATATATATATATAT 87
Qy 2679 cagttccctagtttaagaagtagagcttaccctgcatgtgcatcaagaagaatcagagcac 2738
|||||
Db 86 CAGTTCCGATGTTAAAGTAGAGCTTTACCTGTCAATGTGCATCAGCAAGAAATCATAGGAC 27
Qy 2739 tttaataataaaggtttaaagttttgg 2764
|||||
Db 26 TTTTAAATAAAGGTTTAAAGTTTGG 1

RESULT 4 606 bp mRNA linear EST 18-APR-2001
BG611610
LOCUS
DEFINITION
IMAGE:602613533F1 NIH_MGC_60 Homo sapiens cDNA clone IMAGE:4738833 5',
mRNA sequence.
BG611610
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS
TITLE
NIH-MGC http://mgc.nci.nlm.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
JOURNAL
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: DCTD/DRP
CDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
plate: LCM1609 row: a column: 10
High quality sequence stop: 602.
Location/Qualifiers

FEATURES
Source
1. .606
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4738833"
/clone_1b="NIH_MGC_60"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (T1 phage-resistant)"

Db 33 TGTCTGACA 24

RESULT 6
AA495807/c
LOCUS
DEFINITION AA495807 461 bp mRNA linear EST 11-AUG-1997
3', mRNA sequence.
ACCESSION AA495807.1 GI:2229128
VERSION
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 461)
AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Giesel, G., Jost, S.,
Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B.,
Schellenberg, K., Steptoe, M., Tan, F., Theisling, B., White, Y., Wyllie,
J., Waterston, R. and Wilson, R.
TITLE WashU-Merck EST Project 1997
JOURNAL
COMMENT Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
Seq Consortium (info@image.llnl.gov) for further information.
Seg primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 404.
Location/Qualifiers
1. 461
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:768390"
/clone_lib="Soares_NhHMPu_S1"
/tissue_type="Pooled human melanocyte, fetal heart, and
pregnant uterus"
/lab_host="DH10B"
/note="Organ: mixed (see below); Vector: pT7T3D-Pac
(Pharmacia) with a modified polylinker; Site_1: Not 1;
Site_2: Eco RI; Equal amounts of plasmid DNA from three
normalized libraries (melanocyte 2NbH, pregnant uterus
NbHPU, and fetal heart NbH19W) were mixed, and ss circles
were made in vitro. Following HAP purification, this DNA
was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from pools of
5,000 clones made from the same 3 libraries. The pools
consisted of I.M.A.G.E. clones 260232-265223,
340488-345479, and 484488-489479."

BASE COUNT 155 a 80 c 67 g 159 t

ORIGIN

Query Match 11.3%; Score 404.2; DB 9; Length 461;
Best Local Similarity 97.9%; Pred. No. 3.7e-69;
Matches 428; Conservative 0; Mismatches 7; Indels 2; Gaps 2;

QY 3147 gtgcgttttaagaacaactaaatgtgttaattaccacacttag-atgggtgaatgt 3205
|||||
Db 461 GTGCTGTTTAAGAAAGACCTAAGTGTATATTCACCCACTTTAAGATGGGTAAATGT 402
|||||

QY 3206 tatggtgtgtaaatatctcgttaagacgttaaaagaaagagctggaatgcactga 3265
|||||
Db 401 TATGCTGTGTGAATAATCTCAGTAAAGCAAGTTAAAGAAAAAGAGCGAATGCACGA 342
|||||

QY 3266 ttcaagaaacttaattcaggaaggaaggtctgtatgtacacattcacttaagcaga 3325
|||||
Db 341 TTCAGGAACCTTAATTTTCAGGAAGAAAGTCTGTATGTACACATTTCACTTTAAGCAGAA 282
|||||

QY 3326 aatcttctcaagaagaacttactcttctcttgcagctgcagcgtgagatctaa 3385
|||||
Db 281 AATCTTCTTAAGAAAGACCTTACTTCTTGTGCACTGACACAGTGATCACTAA 222
|||||

QY 3386 cttttaactagtt-gtctctctcagctcctcagcttaagatatttgccttcataa 3444
|||||
Db 221 CTTTAACTAGTGTGTTCTTCTCTAGCTCTAGCTTATTAAGATTTTGTCTTCATNA 162
|||||

QY 3445 tctgaaccttaagcaggaagaagaatgttttcagatagtttcaaatacnccaaat 3504
|||||
Db 161 TGTGAACCTTTAAAGCAGAGAAAGAAATGTTTCAGATAGTTCCAAATACACCAAAAT 102
|||||

QY 3505 gtttcacacacaaatctcgttaacacacataatgcccattatgaatatagtta 3564
|||||
Db 101 GTTGAACACCAAAATCTGGAATCAACCATATGACCTTATTAATATATAGTTGA 42
|||||

QY 3565 tagnttgcctcgaaga 3581
|||
Db 41 TAGATTGTCTCGAAGA 25
|||||

RESULT 7
A1682875/c
LOCUS
DEFINITION A1682875 413 bp mRNA linear EST 17-DEC-1999
wc69d11.x1 NCI_CGAP_Pan1 Homo sapiens cDNA clone IMAGE:2323893 3',
mRNA sequence.
ACCESSION A1682875 GI:4893057
VERSION
KEYWORDS
SOURCE EST.
ORGANISM human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 413)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL
COMMENT Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Life Technologies catalog #: 11548-013
DNA Sequencing by: Washington University Genome Sequencing Center
clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
www-bio.llnl.gov/dbtp/image/image.html
Insert Length: 2559 Std Error: 0.00
Seg primer: -40UP from Gibco
High quality sequence stop: 410.
Location/Qualifiers
1. 413
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2323893"
/clone_lib="NCI_CGAP_Pan1"
/tissue_type="adenocarcinoma"
/lab_host="DH10B"
/note="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.72 kb. Life Technologies catalog #:
11548-013"

BASE COUNT 140 a 70 c 58 g 145 t

ORIGIN

Query Match 10.3%; Score 369.8; DB 9; Length 413;
Best Local Similarity 98.2%; Pred. No. 2.1e-62;
Matches 382; Conservative 0; Mismatches 6; Indels 1; Gaps 1;

QY 3133 gatgggtgaatgttatgtgtgtgtaatatctcagtaagcagcttaaaaggaagaagc 3252
|||||
Db 413 GATGGGTGAATGTTATGTGTGTGAATAATCTCAGTAAAGCAAGTTAAAGAAAAAGAGC 354
|||||

QY 3253 tggatgcacgtatcaggaacttaattcaggaaggaaggtctgtatgtacacatttc 3312
|||||

Db	353	TGGAAATGACACGATTTACAGAACTTAATTTTCAGGAAGAAAGTGCTGATGTAC -CATTTC	295
Oy	3313	actttaagcagaaaaactcttcttcaagaatgaacttaacttctctcttgacgtcagca	3372
Db	294	ACTTTAAACCAAAAAATCTTCTTCACGAAGAACACTTTCCTTCTCTGTGCATGCCACCA	235
Oy	3373	cgtagagatacnaactctttaactagttgtctctctctatgtctatcagttatagatttt	3432
Db	234	CGTGGATACCTACTCTTTTAACATACTTGTCTCTCTAGTCTCTACGATTATAGAAATTT	175
Oy	3433	ttagcttcaataatgynaacctttaagcagagagaagaataatgltttcagatagtttcaa	3492
Db	174	TTTGCTTTATATATGTGAACCTTTAAGCAGGAGGAAGAAATGTTTTCAGATATGTTCAA	115
Oy	3493	tacmccaaaatglttgcacaacaaaactatggaatcmaaccataatgcccattgaa	3552
Db	114	TACACCAAAAATGTTTGAACACAAAAATACGGAATCAACCAATNAAGCACTATTGAA	55
Oy	3553	tatatagctgatatagattgtttcttgc meta 3581	
Db	54	TATATAGTTGTAATGATTTGTTCGAAAA	26

RESULT	8
BFI91498	
LOCUS	BFI91498 567 bp mRNA linear EST 02-NOV-2000
DEFINITION	Z39107 MARC 2P1G Sus scrofa CDNA 5', mRNA sequence.
ACCESSION	BFI91498
VERSION	BFI91498.1 GI:11074867
KEYWORDS	EST.
SOURCE	p1g.
ORGANISM	Sus scrofa

REFERENCE
1 (bases 1 to 567)
Fahrenkrug, S.C., Fieking, B.A., Rohrer, G.A., Smith, T.P.L., Casas, E.,
AUTHORS

TITLE
Design and use of two pooled tissue normalized cDNA libraries for EST discovery in swine

JOURNAL
COMMENT
Unpublished (2000)
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA

Tel: 402 762 4366
 Fax: 402 762 4390
 Email: smithgemail.marc.usda.gov
 Single base sequencing. Bases called and alt-trimmed with phred
 v0.980904.e. Vector identified by cross_match with the -mismatch 18
 and -mismatch 12 options.

PCR PRIMERS
 FORWARD: AGGAACAGCTATGACCAT
 BACKWARD: GTTTCACGTCAGCAG
 plate: 66 row: K column: 22
 Seq primer: ATTTAGTGACACTATTAG.

FEATURES	Location/Qualifiers
source	1. .567

BASE COUNT	ORIGIN
157 a	131 c 148 t

Query Match	9.4%	Score 335.6	DB 10	Length 567
Best Local Similarity	78.8%	Pred. NO. 1e-55		
Matches 438	Conservative	0	Mismatches 114	Indels 4
				Gaps 3

QY	1	ctgtgttgtaactgtctaagtttcagagaaagaatagacacaaacacgtgtagctgtatg	60
Db	15	CTGCTTGTGTGGCTGTAGATTGCAAGACAAATATGCAACACACCCTGAGGCTG-ATG	73
QY	61	cgattccagaaaccaagaccacaatcttgcctacatccattcaatcaagttgctgaagtaag	120
Db	74	CGACTCGGAACCAAGACCAATATATATTCATTTCATATATCAAGTTCTCCAGAGAG	133
QY	121	gaaatgacatctggttctgtctctctctacatcttaatttttgaaatatcttccat	180
Db	134	GAATATGACATCTGGGCCGCTTCATCTACATCTTGATTTTGGAACTATTTTGCCTCT	193
QY	181	gggggtggaacggatgtgaagtgtcctcccttgctacattccctcttggaacatcaaa	240
Db	194	GGGATGTGGCGAGATGTACATGCGCCCTGGCTACTTCCTCCCTGTGGAAATCACCAAG	253
QY	241	tgcttgctccacgtccctgcacatgtaacggtgtgagacactcgsgaaatcaaggccgatg	300
Db	254	TGCTTGCCCTCAGTTCCTTCACTGCAACGCTGTGAGACGACTGTGGGAACCAAGCTGACAG	313
QY	301	gacaactgtgtgtgtgtgttggttgccagtgcaat-tctttccaggctctggaactgtagt	359
Db	314	GACAACCTGTGGAGACAAACAATGATGTGCTATGCAACTTGCACCAATATTTTGGCAATTAC	373
QY	360	gatgaaacatttaacgagtgttcacatcggtttcttcaaatgttgactgaaatgtaactca	419
Db	374	TACAAAAGACCTTCCCTCCATCACTTTTGAAGGCAGACGACATCTGGATGG--TGTCACTTCA	431
QY	420	gtggaacttaagaagaagcttcctcctgcatgtgtccttaagaatatatactgacttcagaa	479
Db	432	GTGGATTTTAATTAAGAAAGCTTCCCTCGATGAGCGCTTCAAAAAGTACCATATGATTTCAAA	491
QY	480	gctggaacctgcaaaaacaataagatatacatccatccatctatgctttcagaagactgaa	539
Db	492	GCTGCTCTGCAAAAACATTAAGATTAGTTCGCGATCTGTATGCTTCCAGGACTGTGA	551
QY	540	tagctactactaactg 555	
Db	552	CAGTCTCACTAAACTG 567	

RESULT	9
A0595136/c	
LOCUS	
DEFINITION	
<hr/>	
A0595136	544 bp DNA linear GSS 08-JUN-1999
HS_5424_B1_E06_T7A RPCT-11 Human Male BAC library Homo sapiens	
genomic clone Plate=1000 Col=11 Kow-J, DNA sequence.	

ACCESSION A0595136 GI:5026722
 VERSION A0595136.1
 KEYWORDS GSS.

SOURCE ORGANISM	human, Homo sapiens
-----------------	------------------------

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE
AUTHORS
1 (bases 1 to 544)
Mahairs,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,

Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and Hood, L.

TITLE	Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
<p> Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome 05/03/2004 11:05:37 0230 0244 410003 </p>	

JOURNAL
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9139-9144 (1999)
MEDLINE
99380589

COMMENT **Contact:** Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center

University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 538 3710

Tel: (206) 616-3618
Fax: (206) 616-3887

Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC

library availability, please contact pieter de jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from

BACpAC Resources (http://bacpac.med.buhalo.edu/ordering_pac.htm)

or from Resear h Genetics (info@resgen.com). BAC end Web Server:
<http://www.htsc.washington.edu>
 Plate: 1000 row: J column: 11
 Seq primer: T7
 Class: BAC ends
 High quality sequence stop: 544.
 Location/Qualifiers
 1..544

FEATURES
 source
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="Plate=1000 Col=11 Row=J"
 /clone_lib="RPCI-11 Human Male BAC Library"
 /sex="male"
 /note="Vector: pBACe3.6; Site.1: EcoRI; Site.2: EcoRI;
 Male blood DNA was isolated from one randomly chosen donor
 and partially digested with a combination of EcoRI and
 EcoRI Methylase. Size selected DNA was cloned into the
 pBACe3.6 vector at EcoRI sites"
 BASE COUNT 163 a 165 c 81 g 128 t 7 others
 ORIGIN

Query Match 7.8%; Score 279.4; DB 12; Length 544;
 Best Local Similarity 81.6%; Pred. No. 1.1e-44;
 Matches 347; Conservative 0; Mismatches 76; Indels 2; Gaps 2;

QY 1388 attattcgcgacgcgagcttgacctaagttcgtgagaaataaataagatcgc 1447
 || |||||
 DB 513 ATATATTATTGATGAGAGTTGGCTTAAGTCTCGTAGG-GTACAAATTAAGCATCGC 455
 QY 1448 agcttgagatgagagatcattgtcagcttgaagatcttggccattctgcacag 1507
 |||||
 DB 454 AGCTGTATGAAGAGATCCATTCAGCTGTAGAGATGTTGGCATTGTCTCCACAG 395
 QY 1508 aagatcagtttactgttaacatttcgaacattggaataacatcgtcattc 1567
 |||||
 DB 394 GAGTATCAGTTGAGTGTACATGTGTGACATGGAAGATGACATGCGTGTATC 335
 QY 1568 cttttagatgtgagagcccggaataagcaaatcaatcgttctcatttga 1627
 |||||
 DB 334 GTTTTAAATGTGAGA-CTTGAATGCGAAGCAATGGAAGTGTCTCATTTTGA 276
 QY 1628 ttactggtttatagtgcttcattcattgagcaataaggaattttcaaaact 1687
 |||||
 DB 275 TTGGGGTTTAAAGTGGCTTTCATTCATTGACAGCAATTAAGATTTCAGAGCTACT 216
 QY 1688 atggacacaatgagatagcttcctccttcattcagaagaatcacagaatgagccc 1747
 |||||
 DB 215 ATGGACCAATGAGTGGCTTCCTCTCATTCAGAAAGATACGGGAGTATTGGAGCCC 156
 QY 1748 agattatcagtgagcaatttctcgtgtatatttggccgcatcattacatagtt 1807
 |||||
 DB 155 AGATTATTATGAGTGACCTGTGTGTGTGGAGACTTAAGGGCTTAATAATGTTGTTGA 96
 QY 1808 ttccc 1812
 |||||
 DB 95 ATTCC 91

RESULT 10
 AI375172/c 286 bp mRNA linear EST 18-MAR-1999
 LOCUS AI375172
 DEFINITION 3', mRNA sequence.
 ACCESSION AI375172
 VERSION AI375172.1 GI:4175162
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 286)
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaps-r@mail.nih.gov
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert Length: 651 Std Error: 0.00
 Seq primer: -400P from Glibco
 High quality sequence stop: 276.
 Location/Qualifiers
 1..286

FEATURES
 source
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2063423"
 /clone_lib="Soares_NhHMPu_S1"
 /tissue_type="Pooled human melanocyte, fetal heart, and
 pregnant uterus"
 /lab_host="DH10B"
 /note="Organ: mixed (see below); Vector: pT73D-Pac
 (Pharmacia) with a modified polylinker; Site.1: Not I;
 Site.2: Eco RI; Equal amounts of plasmid DNA from three
 normalized libraries (melanocyte 2NbH, pregnant uterus
 NbHpu, and fetal heart NbH119W) were mixed, and ss circles
 were made in vitro. Following HAP purification, this DNA
 was used as tracer in a subtractive hybridization
 reaction. The driver was PCR-amplified cDNAs from pools of
 5,000 clones made from the same 3 libraries. The pools
 consisted of 1.M.A.G.E. clones 260232-265223,
 340488-345479, and 484488-489479."

BASE COUNT 81 a 56 c 56 g 93 t
 ORIGIN

Query Match 7.7%; Score 275; DB 9; Length 286;
 Best Local Similarity 99.7%; Pred. No. 7.9e-44;
 Matches 286; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 2136 cagcaaaaggtcagaacatagctcattcattcattcgttggaatgagcattgca 2195
 |||||
 DB 286 CAGCAAAAGGTCAGAAACATATGCTCATCATCTGCGTGAATAATGTGCCACATGCA 227
 QY 2196 ggaagatccacactgaatgaagaagccggaacctttacatacccccgtgaaagtact 2255
 |||||
 DB 226 GGAGATGCCACCTGATTAATGAAGCCGACCTTTTACATACCCCTGTGAATGTACT 167
 QY 2256 gatttcacatcaacagagactcaattcctatcattcagctgactgaaatcattcttc 2315
 |||||
 DB 166 GATTTCATCAATCAACGAGACTCAATTCCTATTCAATGACTGACTGAAATTCATTCTTC 107
 QY 2316 gcagagaatactgtgggggtgtcttcattcagaggtattctgtagtaaaatgaataccaca 2375
 |||||
 DB 106 GCAGAGATACTGTGGGGGTGCTTCATGAGGAGGATTTACTGCTATG-AAATGAATACACCA 48
 QY 2376 aaattaattataataatagtaagataaattttcaagaagactg 2422
 |||||
 DB 47 AAATTAAATTATTAATAGCTAAGATTAATTTTACAAAGGACATG 1

RESULT 11
 BB635922 736 bp mRNA linear EST 26-OCT-2001
 LOCUS BB635922
 DEFINITION BB635922 RIKEN full-length enriched, 0 day neonate thymus Mus
 musculus cDNA clone A430090H24 5', mRNA sequence.
 ACCESSION BB635922
 VERSION BB635922.1 GI:16471894
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 736)
 AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.

TITLE
JOURNAL
COMMENT

Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Arikawa,T., et al. 2001)
Unpublished (2001)
Contact: Yoshinori Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@sc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
Wagl,K., Fujiwaki,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watanabe,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multichannel sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaoka,I., Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.
e mouse tissues.

FEATURES
source

Location/Qualifiers
1. /736
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="A430090H24"
/clone_lib="RIKEN full-length enriched, 0 day neonate thymus"
/tissue_type="thymus"
/dev_stage="0 day neonate"
/lab_host="DH10B"
/note="Site.1: Sali; Site.2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 459.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCGAGTATTAATTAATTAATCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

BASE COUNT
ORIGIN
188 a 191 c 173 g 184 t

Query Match 7.3%; Score 261.4; DB 9; Length 736;
Best Local Similarity 69.0%; Pred. No. 3.7e-41;
Matches 402; Conservative 0; Mismatches 176; Indels 5; Gaps 3;
QY 176 ctaatgggagtgacagagatgttaagtcccttgctatctcccttgctgggaacatca 235
DB 100 CTCCTCGAAGTGCACAGGAGTGTGCTCCCTCGCTCCCTCCCTCGGGAATGA 159
QY 236 caaagtcgtctgaagctccctgcacgtgaaggtgtgagcagctcggaacacagccg 295
DB 160 GCAGGTGCTGCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 219
QY 296 atgaggaacactgt 355
DB 220 ATGAGGACCACTGCGGAGCAACAAATGAGGCGTCTGCTGCTGCTGCTGCTGCTG 279
QY 356 acgtgatgaaccatttaagagatgttccatcggtttcttaaatgtgacgaattgcac 415
DB 280 ATTACATATAACGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 338
QY 416 ttcaatgggaacttaataagaagctccctgcacgtgtgtgtgtgtgtgtgtgtgtgt 475
DB 339 TGCAGAGGAATCTCATAGAACACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 398
QY 476 agaagcttgacactgtcaaaacaaataaataacatcacatccatcattgtcttcaaga 535
DB 399 AGAAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 458
QY 536 tgaatgcttactaactgtatctcagtcatacagaataaccttcgaacccgggtgtg 555
DB 459 TGGCAGCCCTCAATCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 518
QY 596 tttttaagaatcttcaagaactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 655
DB 519 GTTTGAAGATCTCCACAGACCTGGAATGCTGATTAATGAAGAAATCACTGACGCA 578
QY 656 ttcccccacacatttatatgatgaataatctcttattcttcttgaatcgtatgaatga 715
DB 579 TTCCACACCTCA--CAATTAAGGACTTAATTTCTTTTATTTGAGCTGGA--GAAAG 634
QY 716 tctctacccgttactcgtatgaacacctctcgtcaacacatgcg 758
DB 635 CCTCACCCCTGCTGGAACCCCTTGTGCTGCAAGGCCGCC 677

RESULT 12
AA495745 244 bp mRNA linear EST 11-AUG-1997
LOCUS 2w04c04.r1 Soares_NhmPv_S1 Homo sapiens cDNA clone IMAGE:768294
DEFINITION 5', mRNA sequence.
ACCESSION AA495745
VERSION AA495745.1 GI:2229066
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 244)
Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisler,G., Jost,S., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Mattin,J., Moore,B., Schellenberg,K., Stepien,M., Tan,F., Theising,B., White,Y., Wyllie,T., Waterston,R. and Wilson,R.
WashU-Merck EST Project 1997
Unpublished (1997)
Contact: Wilson RK
JOURNAL
COMMENT Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through INMIL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -28m13 rev2 ET from Amersham

High quality sequence stop: 119.

FEATURES
Location/Qualifiers

1. 244
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:768294"
/clone_id="Soares_NbHMPu.S1"
/issue_type="Pooled human melanocyte, fetal heart, and pregnant uterus"
/lab_host="DH10B"
/note="Organ: mixed (see below); Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (melanocyte 2NbHMP, pregnant uterus NbHMP, and fetal heart NbHMP19M) were mixed and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 260232-265223, 340488-345479, and 484488-489479."

BASE COUNT
ORIGIN

78 a 46 c 36 g 84 t

Query Match
Best Local Similarity 100.0%; Pred. No. 9.4e-38; Length 244;

Matches 244; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2762 tggaaatactcagtgattgtcattcattacatagaataatgtctgactgtttgcaaaataatattct 2821
|||||
Db 1 TGGATACTAGTGTATTGTCATCATAGAAATGTCACGTGTTGCAAAATTAATATCT 60

QY 2822 gtttaagaatccatcttaacctctttaaagttccatcacacttggagagccaacaaca 2881
|||||
Db 61 GTTTAAAGATCCATCTTACCTCTTTAAAGTTCCATACACTTGAGAGCAACACACA 120

QY 2882 tattattactcaaaaagtgcttgcttagaactcaaaaagcactcttcttgacatt 2941
|||||
Db 121 TATTATTACTTAAAGATGCTTGCTTGAAGAACTCAAAAGCAGCTCTTTGGCAGCTT 180

QY 2942 cctgcaccagttctcctctgtctttaaagtaacatcatcatatgtaagtaagagag 3001
|||||
Db 181 CCTGCCAGCTTCTCTCTTCTTAAATGAACATCATATGGAATTGGAATAGAGAG 240

QY 3002 tatg 3005
|||||
Db 241 TATG 244

RESULT 13

AA122078 292 bp mRNA linear EST 19-NOV-1996

LOCUS
DEFINITION ZK93910.r1 Soares-pregnant_uterus_NbHMP Homo sapiens cDNA cloneIMAGE:490434 5', mRNA sequence.
AA122078
AA122078.1 GI:1678116ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 292)REFERENCE
AUTHORS
Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman

M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,

Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaaskis,E., Waterston

R., Williams,A., Wohlmann,P. and Wilson,R.
The WashU-Merck EST Project
Unpublished (1995)Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800

Fax: 314 286 1810

Email: est@wustl.wustl.edu

This clone is available royalty-free through LNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 210.FEATURES
Location/Qualifiers

1. 292
/organism="Homo sapiens"
/db_xref="GDB:3805021"
/db_xref="taxon:9606"
/clone="IMAGE:490434"
/clone_id="Soares_pregnant_uterus_NbHMPu"
/sex="Female"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: uterus; Vector: pT7T3-Pac; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', AACGTGAAGAAATGCGCGCCCTTTTCTTTTCTTTTCTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization. Library constructed by M. Fatima Bonaldo."

BASE COUNT
ORIGIN
94 a 52 c 51 g 92 t 3 othersQuery Match
Best Local Similarity 95.9%; Pred. No. 4.8e-37; Length 292;

Matches 278; Conservative 0; Mismatches 8; Indels 4; Gaps 3;

QY 2759 ttgtgaaatactcagtgattgtcattcattacatagaataatgtctgactgtttgcaaaataatattct 2818
|||||
Db 3 TTTTGAAATACAGTGTATTGTCATCATAGAAATGTCACGTGTTGCAAAATTAATAT 62

QY 2819 tctgtttaaagaatccatcttaacctctttaaagttccatcacacttggagagccaaca 2878
|||||
Db 63 TCTGTTTAAAGATCCATCTTACCTCTTTAAAGTTCCATACACTTGAGAGCAACACA 122

QY 2879 acatatattactcaaaaagatgcttgcttagaactcaaaaac--agcactctcttgg 2936
|||||
Db 123 ACATTTTATTACTTAAAGATGCTTGCTTGAAGAACTCAAAAGCAGACTTCTTTTGG 182

QY 2937 caactccgtccagttctc--tttgctttaaagtaacatcatcatatgtaagtaagta 2995
|||||
Db 183 CACTTCTGCGCCAGTCTTCTCTTGTCTTAAATGAACATCATATGGAATGGAATTA 242

QY 2996 gtagagatagagtaac-ggcagagagaagtggatcagaanaactagaatggg 3044
|||||
Db 243 GCAGAGTATGATGATCGGAGAGAGTGCAGAAANCTAATGATGAGG 292

RESULT 14

AO745959/c 786 bp DNA linear GSS 16-JUL-1999

LOCUS
DEFINITION HS_2275_A1-G05_T7C CIT Approved Human Genomic Sperm Library D Homosapiens genomic clone Plate-2275 COI-9 Row-M, DNA sequence.
AO745959
AO745959.1 GI:5523481ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 786)REFERENCE
AUTHORS
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,

Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D., and

Hood,L.
Sequence-tagged connectors: A sequence approach to mapping andscanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)

Thu Sep 5 14:39:12 2002

us-09-647-067-7.rst

Page 11

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 5, 2002, 02:58:00 ; Search time 6304.42 Seconds
(without alignments)
11896.528 Million cell updates/sec

Title: US-09-647-067-7

Perfect score: 3584

Sequence: 1 cgccttgtaactgctaaga.....tagntttgtctgaataacc 3584

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenBml: *
1: gb_ba: *
2: gb_htg: *
3: gb_in: *
4: gb_om: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_sts: *
12: gb_sy: *
13: gb_un: *
14: gb_vi: *
15: em_da: *
16: em_fun: *
17: em_hum: *
18: em_in: *
19: em_mu: *
20: em_om: *
21: em_or: *
22: em_ov: *
23: em_pat: *
24: em_ph: *
25: em_pl: *
26: em_ro: *
27: em_sts: *
28: em_un: *
29: em_vi: *
30: em_htg_hum: *
31: em_htg_inv: *
32: em_htg_other: *
33: em_htgo_inv: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result Query
No. Score Match Length DB ID Description

C	1	1996.4	55.7	2274	9	AF190500	AF190500 Homo sapi
	2	1571.4	43.8	187089	2	AC107219	AC107219 Homo sapi
	3	1189.4	33.2	1191	6	AX128507	AX128507 Sequence
	4	687.8	19.2	2436	10	AF453828	AF453828 Homo sapi
	5	607.6	17.0	2539	10	AF346501	AF346501 Mus muscu
	6	439.2	12.3	474	6	AX088165	AX088165 Sequence
	7	437.2	12.2	1068	6	AX148176	AX148176 Sequence
C	8	247.8	6.9	171000	6	AC098607	AC098607 Rattus no
	9	246.2	6.9	575	5	AX356204	AX356204 Sequence
	10	194.2	5.4	3756	3	LSGPCR	Z23104 L. stagnalis
	11	187.4	5.2	170522	9	AL138708	AL138708 Human DNA
	12	187.4	5.2	176565	2	AL136106	AL136106 Homo sapi
	13	175.4	4.9	171000	2	AC098607	AC098607 Rattus no
	14	173	4.8	125322	2	AC108017	AC108017 Homo sapi
	15	173	4.8	180122	9	AC019341	AC019341 Homo sapi
C	16	162.8	4.5	180673	10	AC068627	AC068627 Mus Muscu
	17	162.8	4.5	213462	10	AC077689	AC077689 Mus Muscu
	18	158.4	4.4	592	6	AX320420	AX320420 Sequence
C	19	138.2	3.9	125322	2	AC108017	AC108017 Homo sapi
	20	114.2	3.2	576	6	AX320477	AX320477 Sequence
C	21	110	3.1	1018	6	AX147820	AX147820 Sequence
	22	103.8	2.9	155163	2	AC098990	AC098990 Rattus no
	23	96	2.7	1015	6	AX254453	AX254453 Sequence
C	24	89.2	2.5	155163	2	AC098990	AC098990 Rattus no
	25	85.8	2.4	156490	10	AL603709	AL603709 Mouse DNA
C	26	81.6	2.3	67157	2	AC101958	AC101958 Mus muscu
	27	76	2.1	10197	1	AE006607	AE006607 Streptoco
C	28	76	2.1	120871	9	AC016605	AC016605 Homo sapi
	29	73.6	2.1	123050	9	AC008917	AC008917 Homo sapi
	30	73.6	2.1	2895	3	DMTARTAN	U02078 Drosophila
	31	73.6	2.1	3192	3	AY051439	AY051439 Drosophila
	32	73.6	2.1	96877	2	AC017776	AC017776 Drosophila
C	33	73.6	2.1	162343	3	AC010659	AC010659 Drosophila
	34	73.6	2.1	271729	3	AE003539	AE003539 Drosophila
	35	72.2	2.0	1080	6	AX254459	AX254459 Sequence
	36	68.2	1.9	2019	6	AX179624	AX179624 Sequence
	37	64.4	1.8	5261	10	AF144627	AF144627 Mus muscu
	38	64.2	1.8	17443	5	AC024806	AC024806 Caenorhab
	39	63.8	1.8	3075	5	AB005587	AB005587 Cynops py
	40	63.2	1.8	5954	9	AB058757	AB058757 Homo sapi
	41	63.2	1.8	8594	9	HSY19205	Y19205 Homo sapien
	42	63.2	1.8	161341	9	HSBG115M3	AL109653 Human DNA
	43	62.4	1.7	4015	9	AB018349	AB018349 Homo sapi
	44	60.4	1.7	1496	3	CEY53C12D	AL033123 Caenorhab
	45	60.4	1.7	2621	9	AB046639	AB046639 Macaca fa

ALIGNMENTS

RESULT 1
AF190500
LOCUS
DEFINITION Homo sapiens leucine-rich repeat-containing G protein-coupled
receptor 7 (UGR7) mRNA, complete cds.
VERSION AF190500
KEYWORDS AF190500.1 GI:10441729
SOURCE
ORANISM human.
human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 2274)
Hsu,S.Y., Kudo,M., Chen,T., Nakabayashi,K., Bhalla,A., van der
Spek,P.J., van Duin,M. and Hsueh,A.J.
The three subfamilies of leucine-rich repeat-containing G
protein-coupled receptors (UGR): Identification of UGR6 and UGR7
and the signaling mechanism for UGR
Mol. Endocrinol. 14 (8), 1257-1271 (2000)
JOURNAL
MEDLINE
PUBMED
2 (bases 1 to 2274)
Hsu,S.Y. and Hsueh,A.J.W.

TITLE Direct Submission
JOURNAL Submitted (28-SEP-1999) GYN/OB, Stanford University, MSOB S385
Stanford, CA 94305, USA
FEATURES Location/Qualifiers

BASE COUNT	669 a	460 c	421 g	724 t
ORIGIN				

Query Match	55.7%;	Score 1996.4;	DB 9;	Length 2274;
Best Local Similarity	94.9%;	Pred. No. 0;		
Matches 2158;	Conservative	0;	Mismatches 11;	Indels 105; Gaps 5

QY	124	atgaacatcgtgttcctgtctctcttcttaacatctaaatcttttgaaaatatttttcctatgg	183
Db	1	ATGACATCTGGTCTGTCTCTCTCTCTACATCTTAATATTTTGGAAATAATTTTCTCATGG	60
QY	184	ggtgacagagatgltcaagtgtcctcttgactattcccttgttggaaacatcacaaagtgc	243
Db	61	GGTGGACAGAGATGTCATGCTGCCCTTCCCTGATTTCCCTGTGGAAATCATCAAAAGTGC	120
QY	244	tgtcctcagctcctcgtcactgtgtaaggtgttgaaagactgcgggaatcagcagatlgagac	303
Db	121	TTGCTCTACAGCTCTCTCACTGTAAACGCTGTGACGATCGGGAAATCAGGCCAGATGAGAC	180
QY	304	aactgtg	310
Db	181	AACGTGGAGACACAAATGATGATGTCATGCATTTTGCAATATTTTGGCAGTTACTAC	240
QY	311	-----tggtgtgtttgtg	323
Db	241	AAATGACTTCCCAATATCTTTTGAGGACAGAAACACCTGATGTMTTGGTGGTCTGTG	300
QY	324	ccagatgc-atgctcttgcc-agctctgagcttgactg-gatgaacca--tttagcagt	378
Db	301	CCAGTGCATATGCTTTGGCAAGGTGTGAGGCTTGACTGTGATGAACCAATTTTACGAGCT	360
QY	379	gttccatcggttctcttccaatgtgactgccaatgtcactctcagttggaactaatagaag	438
Db	361	GTTCATCTCGTTCTTTCAAATGTGACTCAATGTCACTTCAGTGGAACCTTAATAGAAG	420
QY	439	cttccctccgattgtgttccaagaattatcatgatctcttcaagaagcttgcaccttcaaaacaat	498
Db	421	CTTCCCTCCGATTTGCTTCAAGATTAATCATGATCTTTCGAAGCTGTACCTSCAAAACAT	480
QY	499	aagattacatcatcctcatctatagtcttcagaggaactgaaatagacttactaaactgtat	558
Db	481	AAGATTATCATTCATCTCCATCTATCTTTCAGAGCATCGAATACCTTACTAAACTGTAT	540

Qy	559	cccfatcabaacagataaactccctccgaaagccggygtgttttgaagatccttcacgacta	618
Db	541	ctcagatcaataacagataaactccctccgaaagccggygtgttttgaagatccttcacagacta	600
Qy	619	gaatgctatataatgaataataacccctcagtcgaattcccccacacattatga	678
Db	601	gaatgctatataatgaataataacccctcagtcgaattcccccacacattatga	660
Qy	679	ctaaatccttatctcctcagtcgaataacgctcccaaccggttaccgtataa	738
Db	661	ctaaatccttatctcctcagtcgaataacgctcccaaccggttaccgtataa	720
Qy	739	ccctctctcacaacactgcccaagactacgttgctgagaccttgaaggaaccatcacat	798
Db	721	ccctctctcacaacactgcccaagactacgttgctgagaccttgaaggaaccatcacat	780
Qy	799	aattagaatattgactcttattctccgagaaatttaacgtttctagtgatgagaa	858
Db	781	aattagaatattgactcttattctccgagaaatttaacgtttctagtgatgagaa	840
Qy	859	aacaaatlaactlaactlaaattgaanaactctttgcaacctccagaactgagatg	918
Db	841	aacaaatlaactlaactlaaattgaanaactctttgcaacctccagaactgagatg	900
Qy	919	gatttggagaatgaataagtttgaanaacttccacacgcttatccaaagacccgaagag	978
Db	901	gatttggagaatgaataagtttgaanaacttccacacgcttatccaaagacccgaagag	960
Qy	979	ctgtcacaaattgaactcttcccttaataccaatccagaanaattcaagcaaccaattgat	1038
Db	961	ctgtcacaaattgaactcttcccttaataccaatccagaanaattcaagcaaccaattgat	1020
Qy	1039	tactctgtcaactcaagctctccacacgttagaaggaattgaattccaatatccaa	1098
Db	1021	tactctgtcaactcaagctctccacacgttagaaggaattgaattccaatatccaa	1080
Qy	1099	aggatcttgaacctctatgaatactctccacatatatttaagaanaattccagtaactgt	1158
Db	1081	aggatcttgaacctctatgaatactctccacatatatttaagaanaattccagtaactgt	1140
Qy	1159	gggatgacacacatgttcgagctgtgaaccaaacaacatgagtaatttcactcaga	1218
Db	1141	gggatgacacacatgttcgagctgtgaaccaaacaacatgagtaatttcactcaga	1200
Qy	1219	aactccttggcaagatattccagagagatactgtctcgggttcatctgcaagtaccctgc	1278
Db	1201	aactccttggcaagatattccagagagatactgtctcgggttcatctgcaagtaccctgc	1260
Qy	1279	tttggaaacatlttgtcatattgcatatgagacacttatatacagctctggaacaaagctgat	1338
Db	1261	tttggaaacatlttgtcatattgcatatgagacacttatatacagctctggaacaaagctgat	1320
Qy	1339	ggcatgtgcaactcttctcctgctgtgacgactgtcaattggaatataattatctgtg	1398
Db	1321	ggcatgtgcaactcttctcctgctgtgacgactgtcaattggaatataattatctgtg	1380
Qy	1399	atcggagacttgaactaaagtttctgtggagaataacataagcatgcgaactgttgatg	1458
Db	1381	atcggagacttgaactaaagtttctgtggagaataacataagcatgcgaactgttgatg	1440
Qy	1459	gagagatcactatgtcaagctttgtaggtcttggccaattgtccacagaaagatcaagct	1518
Db	1441	gagagatcactatgtcaagctttgtaggtcttggccaattgtgtccacagaaagatcaagct	1500
Qy	1519	ttactgttaacatcttcgaacttggaaanaatacctgcgaactgttctatccctttagatgt	1578
Db	1501	ttactgttaacatcttcgaacttggaaanaatacctgcgaactgttctatccctttagatgt	1560
Qy	1579	gtgagacctgggaaatgacagaaacaaattacagttctgattctcatcttgatattacggttt	1638
Db	1561	gtgagacctgggaaatgacagaaacaaattacagttctgattctcatcttgatattacggttt	1620

```
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
```

```

Contact: submissions@watson.wsl.i.edu
Project Information
Center project name: H_NH0575B04

----- Summary Statistics -----
Sequencing vector: M3; 0%
Sequencing vector: plasmid; 100%
Chemistry: Dye-primed ET; 0% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 182907 bases at least Q40
Consensus quality: 183613 bases at least Q30
Consensus quality: 184039 bases at least Q20
Insert size: 200000; agarose-fp
Insert size: 186189; sum-of-contigs
Quality coverage: 6.48 in Q20 bases; agarose-fp
Quality coverage: 7.47 in Q20 bases; sum-of-contigs
-----

* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 3402: contig of 3402 bp in length
* 3403 3502: gap of unknown length
* 3503 10112: contig of 6610 bp in length
* 10113 10212: gap of unknown length
* 10213 23531: contig of 13319 bp in length
* 23532 23631: gap of unknown length
* 23632 34706: contig of 11075 bp in length
* 34707 34806: gap of unknown length
* 34807 53512: contig of 18706 bp in length
* 53513 53612: gap of unknown length
* 53613 113083: contig of 59471 bp in length
* 113084 113183: gap of unknown length
* 113184 175773: contig of 62590 bp in length
* 175774 175873: gap of unknown length
* 175874 178033: contig of 2160 bp in length
* 178034 178133: gap of unknown length
* 178134 182702: contig of 4569 bp in length
* 182703 182802: gap of unknown length
* 182803 187089: contig of 4287 bp in length.
*
* Location/Qualifiers
*
* 1.187089
* /organism="Homo sapiens"
* /db_xref="taxon:9606"
* /chromosome="4"
* /clone="RP11-575B4"
*
* 1. 3402
* /note="assembly_name:Contig10"
* 3503. 10112
* /note="assembly_name:Contig11"
* 10213. 23531
* /note="assembly_name:Contig12"
* 23632. 34706
* /note="assembly_name:Contig13"
* 34807. 53512
* /note="assembly_name:Contig14
* vector_side:right"
* clone_end:SP6
* vector_side:right"
* 53613. 113083
* /note="assembly_name:Contig15
* clone_end:T7
* vector_side:right"
* 113184. 175773
* /note="assembly_name:Contig16"
* 175874. 178033
* /note="assembly_name:Contig17"
* 178134. 182702
* /note="assembly_name:Contig18"

```

```
misc_feature      182803..187089
                    /note="assembly_name:Contig9"
BASE COUNT      54898 a 36771 c 37629 g 56880 t      911 others
ORIGIN
Query Match      43.8%; Score 1571.4; DB 2; Length 187089;
Best Local Similarity 98.4%; Pred. No. 0;
Matches 1595; Conservative 0; Mismatches 25; Indels 1; Gaps 1;

QY 1961 aattcttcacgttcaggtgtagaataaccaggtaccataacctctggtagtatt 2020
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 37082 AACACATAAATGATTCATTCCTTTCTTTTGTAGTACCAATACCTCTGGGTGATGATTT 37023

QY 2021 ttattctgccattacaagtcgtttgaacccaattctcttactctgaccacaagaccat 2080
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 37022 TTATTCTGCCCATTTACAGTGCTTTGAAACCAATTCTTACTCTTACCCACAAAGACCAT 36963

QY 2081 ttaagaagaatgatcaccggttttggtataactacagacaagaanaaactatggacagca 2140
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 36962 TTAAAGAAATGATTCATTCGGTTTGGTATTAACACAGAAAGAAATCTATGACAGACA 36903

QY 2141 aaggtcagaagaacataatgcccacatcatcctgggtggaagtggtgccactcagagaga 2200
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 36902 AAGGTGAGAAACATATGCTCCATCATCTGGGTGGAATGTGGCCACTGACAGAGA 36843

QY 2201 tgcaccctgagttatgaagccgaccctttccacataccctcgtgaaatgtcaactgatt 2260
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 36842 TGGCACCCTGAGTTAATGAACCCGACCTTTTACATACCCCTGGAATGTCACTGATTT 36783

QY 2261 ctaaatcaacgagactcaattcctatcatgacgtgactctgaaatcaattctctcgaga 2320
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 36782 CTCATATCAACGAGACTCAATTCCTATTCATGACATGACTGAAATTCATTTCTTCGAGA 36723

QY 2321 gaattacgtgggttggtttatgaaggatttactcgtatgaanaatgatacccaaat 2380
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 36722 GAATACGTGGGGGTGCTTCTATGAGGATTTACTGCTGATG- AATGTGAATGCCCAAAATTT 36664

QY 2381 aatttaataatagactgaagataatacttacaagacataggaanaaaataaaatgac 2440
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 36663 AATTATTAATAATAGCTAATGATAATATTTTACAGAGACATGAGGAGAAATTAATAATGAC 36604

QY 2441 taatgctcttaacaaagggaatataatcaataatgatatatactagtagaacattt 2500
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 36603 TAATGCTCTTACAAAGGAGGATATATATCAATATATATATATATAGTACATTTT 36544

QY 2501 gcataaagaatgaagaagaacttaactcagtaacatcatcattcttctaaacgtcat 2560
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 36543 GCATTAAGAAATTAAGAGAAATCTACTGATGATGATGATGATGATGATGATGATGATG 36484

QY 2561 tatgagtaaccactactactgtagcatagcaatgcaatagtcctgggaagtagaagctgc 2620
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 36483 TAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 36424

QY 2621 agaaccttcaatctgtagatgtagttaaagacaaagaactatacaagaatcctcgca 2680
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 36423 AGAACCTTCAATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 36364

QY 2681 gtccctagtttaagaatgaagaacttaactctgcatgtagtaagaagaagatatagaacct 2740
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 36363 GTTCTCTAGTTTAAAGTAGAGACTTACTCTGATGATGATGATGATGATGATGATGATG 36304

QY 2741 ttaataaaggtttaaaggtttggaatactcagtgtaattgcatcataganaaatgctcga 2800
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 36303 TTAAATTAAGGTTTAAAGGTTTGGAAATGATGATGATGATGATGATGATGATGATGATG 36244

QY 2801 ctgtttgcaaaataataatctgttttaagaatcactcctcctcttaagtctcata 2860
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 36243 CTGTTTGCAAAATTAATATCTGTTTAAAGATGATGATGATGATGATGATGATGATGATG 36184

QY 2861 caatttagagacaacacacataatattacttaaaagatgctcttgtagaaactcaaa 2920
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 36183 CACTTGAGACCAACACATATATTATTACTAATAAAGATGCTTGTCTAGAAACTCAAA 36124
```

```
QY 2921 acagacactctcttggcacttccctgcccagtttctctcttctttaaataacatcatca 2980
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 36123 ACAGCACTTCTTTGGGACACTTCGCCGACATTTTCTCTTGGCTTTAATGATACATCATCA 36064

QY 2981 tatggaattggaatgagagagatgtagtacgcgacagaagaatgtagcaagaanaactagaat 3040
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 36063 TATGGAATTTGGAATGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 36004

QY 3041 gagagataacatttactcttgtaggaactctcgaataaactcttctgtagtgaatgaac 3100
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 36003 GAGGATTAACATTTACTACTGATGGAACCTCTGAAATTAATCCTGTATTGTACAGTTAAC 30944

QY 3101 tgatttccaagaagatgccaagaacaaaggcttcttcaacaaacgctgctgttttaaga 3160
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 35943 TGATTTCAACAAGATGACCAAGACAAAGAAAGGCTTTTCAACAACCGTGTGTTTAGA 30884

QY 3161 acagacctaagtggtttaattcacccaacttagatgggtgaatgtagtggtagtgaat 3220
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 35883 ACAGACCTAAGTGTGTTAATTCACCCACTTATGATGGGTGAATGTTATGCTGTGAAT 35824

QY 3221 atctcagtaagaagttaaagaaagaaagcctggaatgacgtattcagaacttaatt 3280
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 35823 ATCTCACTAAAGCATTTAAAGGAAAGAGCTGGAATGCACTGATTCAGAGAACTTAATT 35764

QY 3281 tcaagaagaagaagctgtgtagtacaacttcaactttaaagaagaanaacttcttccaaga 3340
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 35763 TCAGGAAGGAAGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 35704

QY 3341 aatgaacttaacttctcttccagtcacgcgaagtagtagataacttcttaactagttg 3400
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 35703 AATGACATTTACTTCTTCTTGGCACTGCCACACGTGGAATCAATCTTTTAACTTAATTG 33644

QY 3401 ttcttctcagtcctcagcttatgaagatttttgcttcaataatgtagaacttaagc 3460
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 35643 TTCTTCTCTAGTCTCTACGATTAATGAAATTTTTCCTTTAATATGGAACCTTTAAGC 35584

QY 3461 aggaagaagaagaatgcttccagatagtttcaaatacccaanaaalytttgcaacaaaaa 3520
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 35583 AGGAGAAGAAATGTTTTCAGATAGTTTCAATATACACCAAAATGTTGAACACAAATAA 35524

QY 3521 tactgagatcaacacataaagcccttatgaaatataatgtagttagttagttagttag 3580
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 35523 TACTGGAATCAACCAATTAATGCACTTATGATATATATATATATATATATATATATAT 35464

QY 3581 a 3581
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 35463 A 35463

RESULT 3
AX128507 1191 bp DNA linear PAT 15-MAY-2001
LOCUS
DEFINITION
ACCESSION AX128507
VERSION AX128507.1 GI:14134974
KEYWORDS
SOURCE
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1191)
AUTHORS Vogell, G., Wood, L.S. and Merchant, K.
TITLE G protein-coupled receptors expressed in brain
JOURNAL Patient: WO 0131014-A 15 03-MAY-2001;
PHARMACIA & UPJOHN COMPANY (US)
FEATURES
source
1..1191
Location/Qualifiers
1..1191
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..1191
/note="unnamed protein product"
/codon_start=1
```

/protein_id="CAC8938.1"
/db_xref="GI:14134975"
/translation="MFRPLVNLSHITFKFKFYQYCGAPHYRSCPNPDGSSLENLILAS
LIDRFEVWVWVAATCFGNIFVICMRPIYISENLVYMSIISLCCACDLGILYLFVIGG
FDLKEFGEYNKHAQIMESTHCOLVSLILSTREVSULLFLEKYLICIVFPCV
RQKCFRTIVLILMITGFLVAFIPLSKNEFENYCGTGVCPPLSEDEPISGACIY
SVAFEGINLAARIIIVFSGSMFVSHOASITAEIRNOVKEMILARREFIVTID
ALCMIPFVVKELSLIQVEIRPGITSMVYFIPLPISALNPILYITTRFRKMHIRF
WNYRQRKSKMSDKQKTIAPDSFIWEMMPLQEMPELMDLFTYPCEMSLLISQSTRL
NSYS"
BASE COUNT 340 a 229 c 226 g 396 t
ORIGIN
Query Match 33.2%; Score 1189.4; DB 6; Length 1191;
Best Local Similarity 99.9%; Pred. No. 1.2e-232;
Matches 1190; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1102 atgttagactcttaatactctctacatatatttaagaataaccagttactgtgg 1161
|||||
Db 1 ATGTTTAGACCTCTTGAAATCTCTCACATATTTTAAAGAAATTCAGTACTGTGG 60
QY 1162 tatgacacacatgtcgacagctgtaaacacacatgtaattcactctagaagat 1221
|||||
Db 61 TATGACACACATGTTCCAGGTGTAACCAACCTGATGGAATTCATCTAGAGAA 120
QY 1222 ctcttggaagcattatctcagaagaatattgtctgggttctatctcagttactgt 1281
|||||
Db 121 CTCCTGGCAAGCATTTATGAGAGATTTGTCTGGGTTGATCTGCAAGTTACCTGCTT 180
QY 1282 ggaacattttgtcatttgcagcactatatacagttctggaacagctgtatgac 1341
|||||
Db 181 GGAACATTTTGTGATTTGTCATGTCAGACCTTATATCAAGCTGAGAACAGCTTATGCC 240
QY 1342 atgtcaatcattctctctgtctgtgcacagcttaattggaatataattactgtatc 1401
|||||
Db 241 ATGCATCATATTTCTCTGCTGTCGCGACGCTTAATGGAATATATTTATTTGCTGATC 300
QY 1402 ggaagccttgacctaaggttcgtggaagaataacagcagctgtgtgagtgag 1461
|||||
Db 301 GGAGGCTTTTAACTTAAAGTTTCGTGAGAAATACATATAGCATGCGAGCTGTGATGAG 360
QY 1462 agtaccattgtcagctgtgagaagcttggccattctgtccacagaagtaacagttta 1521
|||||
Db 361 AGTACTATGTGTCAGCTTGAAGATCTTTGGCCATTCTGTCCACAGAAATGATGTTTA 420
QY 1522 ctgttaacattctgcacatggaataataacatctgtcatctctttagatgtgtg 1581
|||||
Db 421 CTGTTACATTTTGCACATTGGAATAATACATGCAATTCATTCCTTTAGATGTGTG 480
QY 1582 agacctggaataatgagaacaaataacagttctcattctgattactgtgttata 1641
|||||
Db 481 AGACCTGGAATAATGCAAAATTAAGTTCTGATCTCATTTGATGATACGTTTATA 540
QY 1642 gtgaccttcattccattgagcaataagaaatttcaaaaaactactgtgcacacatgga 1701
|||||
Db 541 GTGGCTTCAATTCATGAGCAATTAAGAAATTTTCAAAAACTCTATGCGACCAATGGA 600
QY 1702 gtagtctccctctcatctcaagaagatacagaagaatttgagccagattactcagtg 1761
|||||
Db 601 GTATGCTTCCCTTTCATTCAGAAATACAGAAATGTTTGAGGCCAGATTTATTCAAGTG 660
QY 1762 gcaattttctgttataatttggccgcaatttcatcaatagtttttctctatgaagc 1821
|||||
Db 661 GCAATTTTCTGTGATTAATTTTGCCGCAATTATACATAGTTTTCCTATGGAAGC 720
QY 1822 atgtttatagtgtcatcaaatgccaataaacaacacatgaaatacaggaatcaagttaaa 1881
|||||
Db 721 ATGTTTATAGTGTTCATCAAAAGTGCATTAACAGCACTGAAATACGGAATCAAGTTAAA 780
QY 1882 aaagaagatgactctccaaaagtttctcttataagattactatgacatgtgctgg 1941
|||||
Db 781 AAAGAGATGATCCTTGCCAAACGTTTTTCTTTATGATTTACTGATGATGATGTCGCG 840

QY 1942 ataccattttgtagtgaaatttcttccactgttcaggttagaataaccaggttacata 2001
|||||
Db 841 ATACCCATTTTTGTAGGAATTTCTTTCACCTGCTTAGGTAGAAATACAGGATACATA 900
QY 2002 acccttggttagtgattttatctcgcattaaacagtgcttgaacccaattctctat 2061
|||||
Db 901 ACCCTTGGGTAGTGAATTTTATTTCTCCATTACAGTGCTTTGAACCCAAATTTCTAT 960
QY 2062 actctgaccacaagaccatttaagaagaattatcactcgttttggtaactacaagaaa 2121
|||||
Db 961 ACTGTACCAACAAGACCATTTAAAGAAATGATTCATGCTGTTTGTGATTACTACAGACAA 2020
QY 2122 agaaaactctatgacagcaaaaggtcagaacacatatgtctcattcattcctgtgtaa 2181
|||||
Db 1021 AGAAAATCTATGACACAAAGGTCAGAAACATATGCTCATCATTCATCTGGGTGAA 1080
QY 2182 atgtgacactgacagagatgacacactgtagttaatgaagccggaacctttcacatccc 2241
|||||
Db 1081 ATGTGCCCATGCGAGAGATGCCACCTGATTAATGAAGCCGACCTTTTCACATACCCC 1140
QY 2242 tgtgaatgtcactgatttctcaatcaacagagactcaattcctatcatga 2292
|||||
Db 1141 TGTGAATGTCACTGATTTTCATCAATCAACGAGACTCAATTCCTATTCATGA 1191
RESULT 4
AF453828
LOCUS 2436 bp mRNA linear PRI 04-FEB-2002
DEFINITION Homo sapiens G protein-coupled receptor affecting testicular
descent (GREAT) mRNA, complete cds.
ACCESSION AF453828
VERSION AF453828.1 GI:18483167
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 2436)
Gorlov,I.P., Kamet,A., Jones,E., Lamb,D., Truong,A., Bogatcheva,N.,
Bishop,C.E., McEleavey,K. and Agoulink,A.I.
Mutations of the GREAT gene cause cryptorchidism
TITLE Unpublished
JOURNAL 2 (bases 1 to 2436)
AUTHORS Agoulink,A.I.
REFERENCE Direct Submission
Submitted (30-NOV-2001) Ob/Gyn, Baylor College of Medicine, 6550
Fannin St., Su. 861, Houston, TX 77030, USA
FEATURES
source
1..2436
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="13"
/map="13q12-q13"
1..2436
/gene="GREAT"
71..2335
/gene="GREAT"
/codon_start=1
/product="G protein-coupled receptor affecting testicular
descent"
/protein_id="AF453828.1"
/db_xref="GI:18483168"
/translation="MIVFLFKFLSRLITMFFLLFYLIVNKPALTOGSMTPS
CQGYPPCGMLITCLPRAFECDKDCDGNADENCGDVGAWTITGVHGNANSVAL
TQECFLKQYPOCCDKETELCEVNDLKVSVNNTVLLSLKRNKIHSPDKVFTKY
TKLKIIFLOHNCIRHISRAKFAFGICNDLILYHNHCITFLRPIFPDLQIOLWLLID
NPTLRISORLEFGLNPFSLSVNNTLEALPKOMCOMPLNVDENGRIRKILKTS
FLSCDSLVYFLFRNOIGVPEPSSIKNIGELDISNTITPESHLERDKLLOK
NLSNPLMTIHRKQFESLKOLQSLDERIEIPINIRNMPKNSHIYFKNRYCTSY
APHVRIOMPLTIDGIISSFEDLLANNILIRIFWVAITFCRNLFIQMRSTIAENTTH
AMSIKILCCADCLMGVYLFPEVGIFDIKRYQXKVALLNWESVQCRIMGLAMLSTEV
SVLLTYLTLERFLVIVFPSPNIRPKRQSVILICIMWAGFLIAVLPWNKDYFGNF

ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS	Overbeek, P.A., Gorlov, I.P., Sutherland, R.W., Houston, J.B., Harrison, W.R., Boettger-Tong, H.L., Bishop, C.E. and Agoulnik, A.I.
TITLE	A transgenic insertion causing cryptorchidism in mice
JOURNAL	Genetics 30 (1), 26-35 (2001)
MEDLINE	21250990
PUBMED	11353515
REFERENCE	2 (bases 1 to 2539)
AUTHORS	Agoulnik, A.I.
TITLE	Direct Submission
JOURNAL	Submitted (06-FEB-2001) Db/Gyn, Baylor College of Medicine, 6550 Fannin Str. Su. 861, Houston, TX 77030, USA
FEATURES	location/Qualifiers
SOURCE	1..2539
gene	/organism="Mus musculus"
CDS	/strain="C57BL/6"
	/db_xref="taxon:10090"
	/chromosome="5"
	1..2539
	/gene="Great"
	194..2407
	/note="Great"
	/codon_start=1
	/product="G protein coupled receptor affecting testicular descent"
	/protein_id="AI08943.1"
	/db_xref="GI:15811373"
	/translation="MMLILHYILLTEVKDFALADSSNVAPLCPKGPFPCCNLTKLPLR AFHGDYDCNGANDENCGDTSGMTTIFGVHNVNKVLTLOCFISQYOHCYCR NELECYKADLKAVPKVSNVTLLSLKKNHRLRVKVESRTLEKYLIONCLTHIS RAEFLAGHNIOLIIYLSHNCITSLSRPGFELKHOLAHAILIDNPTRTSOKSEPMINSI FELPMSNRDLAEPETLCQAQMPLNMWDLANGLIKYTINSFTFLCDSTLYFLPRNDJ GFPEPFSSLNKLGEIDLSSNMATKLPHLFSDHLIQKLNTSSNPLLIVHKNOFGS LKDOLSDLERIEIIPNISGTGMOPMKNSHIYLTFRCSVPVPIKCMPTDISISS LESLANGILRYSVVWIAFIICVGNFLVANSILKAETHAMSIKILCADCKMGVYI LESVGVDIKRGQOYKALLMESVSCORLFLATLSTESVSLLFLTELKEFLVITY PPESNLRIKRCOTVALASTMYVEFLIAAPGETREDYFGNRYGRNGCVPLPHYDAEL FGSGGYSLGIFLGYNLAFLVIYSTYTPMCSIHTRALQTAEVSHISEKVAANKRE FIVEFDAICWLPVFVKIRLSLQVEIPEGTTISLVTFEFLPVNSALNPILTYLTFSPF DKTLQHLHKHRKRPIFYVKKSLSASIVTWDESSLKLGVLISKIALGDSIMRPVP"
BASE COUNT	639 a 641 c 574 g 685 t
ORIGIN	
Query Match	17.0%, Score 607.6; DB 10; Length 2539;
Best Local Similarity	59.8%; Pred. No. 6.5e-114;
Matches 1037; Conservative	0; Mismatches 694; Indels 3; Gaps 1,
Dn	378 ttgtccatcggtttcttcaaatgtagtcgcgaatgtaacatgtaacattcaataaga 437
Dd	535 TGTGGCAAAAGGTTTCACGACACGTAACATTACTCTTAAAGAAAACAATAATCCACAG 594
Oy	438 gottccctcgatggttcgaagaattcatgatattccaagaagcggaccgtcaaaaac 497
Dd	595 ACTTGCAATCAAGGCTTTCACGACAGTATACACAGAAGCTCAGAAAGATATACCTTCAGCACCA 654
Oy	498 taagattacatccatccatctatgcttcagagagactgaatagacctactactaaactgta 557
Dd	655 CTGCATCACACACATCTCCAGGAGAGCATTCCTTGATTAATATCTTAACAATATCTGTA 714
Oy	558 tcctcagtataacagaataacaccttcctcyaaagccgggtgtttttaagaatctcacagact 617
Dd	715 TCTCAGGCATATACCTCAATCTTACTCTCTCAGGCGCTGGGAATATCAAAGACTTCATCAGCT 774
Oy	618 agaagagctgataatctgaagaataaacactagtcgaatttccccacaacaactttatgg 677
Dd	775 TGCTTGCGTAATTTTAGATGACAAACCCGATCCAGAAATCTCAGAAATCTCTTATGgg 834
Oy	678 actaaattctctatctcttagctcctgtagataaagtgctccaccocgttacctgataa 737

Db	835	GTAAACCTCTTTTCTTTTCTTGCCCAATGGGTGGGTAAACCGCGCTCGAGGCCCTTCGTG	1891
Qy	738	acccctctgcacacatgccaagaatcatatggctgcgccttgaaggcaacatcatcca	797
Db	892	AACATTGTGTGCTCAGATGGCTCTCACTCAACTGACCTGGGTGGATCTGGCAACAACTGGAATTAA	951
Qy	798	taatttaagaanaattgacttatttctctgcagtaattaaactgitttagtgaatgagaa	857
Db	952	GTACATTAAGAACATCCACCTTCCTTAACGTCGACGTCGCTCAGCGTTCTGTTTCTGCTAG	1011
Qy	858	aaacaaattaactcaacttaaaatgaanaatacctttgcaectctccgaacactgatatgaat	917
Db	1012	AAATCAAAATTGGTTTGGTTCCAGAGAAAGCAATTTTCTTATTAATAAAATTTAAGGAACT	1071
Qy	918	ggatttggagaagaataaagaattgaaatcttccaccgcttatattccaagaccbgaagaa	977
Db	1072	GGACCTGTCTCAGCAATATGTATTAACAAAACTCCCAAGTCCACCTTTTCAGCGACCTTCACT	1131
Qy	978	gctgcacaattgaatcttctctataatccaaatccagaanaatticaaganaaccaatttga	1037
Db	1132	TCTCCAAACCTGAACCTGTCACTCAACCCCTCTGATGTCTCCAAAGAAACAGATTGG	1191
Qy	1038	ttactctgtcaaacactcaagtctctcgccttagaaggatgtgaatttcaatatccaaca	1097
Db	1192	AAGCTCTAAACAACTTCAGCTCTCTAGACCTGGAAAGATAGATATCCAAATATAGCAC	1251
Qy	1098	aaggatgtttagaccctctcttgatctctctacatatattttaaagaattccagtagt	1157
Db	1252	AGGAATTTCTCAGCCCAATGAAGAACTTTCTCACTATTTTGGAAACCTTTTCGATACTG	1311
Qy	1158	tgggtatgcaccacatgctgcagctgtlaaaccaaacatgataggaaattcatctctaga	1217
Db	1312	CTCCTATGTCCCCCATGTCCGATCTGTATGCCGTGACATGAGATTTCTTCGTCTGA	1371
Qy	1218	gaatctcttggcaagcatatttcaagagagatattgtctggtgttatactgtcagtagctg	1277
Db	1372	GGACCTTTGGCTAAACGGTATCCTCAGAGTGTGTCTGGTTAATAGCTTTTCATTACCTG	1431
Qy	1278	ctttggaaaacttttcttcaatttgcagtcgacatatacaagtccttggaaaacagctga	1337
Db	1432	CGTTGGGAATTTCCCTGTCTCAATAGCCGTGATCTCATTTAAGGCTGAGAAATACACTCA	1491
Qy	1338	tgccatgtcaatcatcttctctcgtctgcgttgcagactgttaaagggaataataatttctgt	1397
Db	1492	CGCATGTCCATCAAAATCCTTTGTGTGCGCATTCGTGATGGGGGTGTACCTGTCTC	1551
Qy	1398	gactcgaagcttgcacctaaagtttcgttggagaatacaataagcatgcgcgtgtgat	1457
Db	1552	CGTGGGGGTCTTTGACATCAATGACGAGGGCGAGTATCGAAGATATGCGCTGCTGTGAT	1611
Qy	1458	ggaagaatraccatgttcagtttgaagatcttggacatcttgcatactgtctcaagaagaatcagt	1517
Db	1612	GGAAAGTGTCCCGCCGCTGCTGTGGCTTCCGTGGCCACAGCTGTCCACGAGAAGTTCGGT	1671
Qy	1518	tttaactgttaacaattctgcacattggaaaaatacacatctgcatgtctatacctcttagatg	1577
Db	1672	GCCTGCTGTGACATCTCTGACGCTGGAGAAAGTTCCTGTGCATAGTATTCCTTTTACGAA	1731
Qy	1578	tgtgaagcttggaaaatgcagaaacaattacagttctgattctcatatttggataactggttt	1637
Db	1732	CGTGCCTGTGGGCAAGCGCCAGACTGCTGTGGCCCTCCGACACATCTGGTGTGGGATT	1791
Qy	1638	tatatgtcttcatcttcatctatgtagcaataaagaatttctcaaaaactactatagcaaca	1697
Db	1792	TCTCATATGCGCGCTTCCGTTTCCAGAGAGGATATTTTTCGGCACTTTTATGGGAAAAA	1851
Qy	1698	tggagatgtctccctctcttcaatccagaagaataagaagaatgttggagccagattatctc	1757
Db	1852	TGGAGTCTGCTTCCCACTTATTTATGACCAAGCAAGAAAGATTTTGGAAAGTAAAGGGTACTC	1911
Qy	1758	agtgcgaatttcttctgttaattatcttggccgattatcatcatagatttcttccatagc	1817

Db 1912 CCTGGATTTTCTAGGTGATGACCTGCTGCTTCCCTGTCATGCGATTTCTGATCT 1971
Qy 1818 aagcattgtttatagttgttcacaaagtgccaaagcgaactgaataacgaatcaagt 1877
Db 1972 CACCAATGTTCTCTCATTAACAAACCCCTTCAAGCTGCAAGAGAGAGACACAT 2031
Qy 1878 taaaaagagatgacatcttgccaaacgttttcttataagattactgaatgacatag 1937
Db 2032 CGGGAAGAGGAGGCGTGTGCAACCGGTTCTTTTATCGTGTCTGATGCCATCTG 2091
Qy 1938 ctgataaccatttctgtatgtaaatcttctcaactgctcaggtagaataaccagttac 1997
Db 2092 CTGGATCCCTGCTGTGTGCTTAAGATCTGCTCTCTCAAGTGAATACCAAGCAC 2151
Qy 1998 cataactcttggtgagttatttcttctgcccataacagttgacgttgaacccaattct 2057
Db 2152 AATCACTTCTGAGATCGTGTGTTTCTTCTTCCGCGGAACAGCGCTTAACCCCATCT 2211
Qy 2058 ctatactctgacacaaagacacatttaagaagaatgacatcggtttgtatataa 2111
Db 2212 CTACACTCTGAGACCTCTTCTTTTAAGACAAGTTGAACAGTTGCTGCACAA 2265

RESULT 6
AX088165 474 bp DNA linear PAT 17-MAR-2001
LOCUS AX088165 Sequence 1 from Patent WO0114548.
DEFINITION AX088165
ACCESSION AX088165.1 GI:13397078
VERSION AX088165.1
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 474)
AUTHORS Duecker,K.
TITLE New g-protein coupled receptor and dna sequences thereof
JOURNAL Patent: WO 0114548-A 1 01-MAR-2001;
MERCK PATENT GmbH (DE)
FEATURES
source location/Qualifiers
1..474
/organism="Homo sapiens"
/db_xref="taxon:9606"
CDS
1..474
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC34686.1"
/db_xref="GI:13397079"
/translation="AOISVAIFLGINLAFTIIVPSYMSFVSHQSAITAPETPNO
VKEMILARFEFFIVTDALCWPPIFVARPLSLDOVEIPGTTISWYVIGYSAINSLN
PIIYTLTPPEKEMHREWHNYRORSMDSKGRHMLHSSGSGCHRCRLS"
BASE COUNT 144 a 92 c 88 g 150 t
ORIGIN

Query Match 12.3%; Score 439.2; DB 6; Length 474;
Best Local Similarity 96.8%; Pred. No. 1.7e-79;
Matches 459; Conservative 0; Mismatches 13; Indels 2; Gaps 1;

Qy 1744 gcccaagattatccagtggaattttctgtgattattgtggcgcattatcatata 1803
Db 1 GCCCAGATTATTACAGGCGCAATTTTCTGGATTATTAATTTGGCGCATTTATCATATA 60
Qy 1804 gtttttccatggaacgacgttttattagtttcatcaaaagtgccaaagcaactgaa 1863
Db 61 GTTTTTCCTATGGAACATGTTTATAGTTCATCAAAAGTGCCATTAACGAAACGAA 120
Qy 1864 atacggaatcaagttaaaaaagagatgacatcttgccaaacgttttcttataagattc 1923
Db 121 ATACGGAATCAAGTTAAAAAAGAGATGATCTTGCCAAAGCTTTTCTTATAGTATTT 180
Qy 1924 acgatgcatatgctgataccattttgttgtaaatcttctacagctcaagta 1983
|||||

Db 191 ACTGATGATTATGCTGATACCCATTTTGTAGGGAACCCCTTCACTGCTCAGGTA 240
Qy 1984 gaataaccaggtacataaccctctggttagtgatttattctgcccattaaacagttct 2043
Db 241 GAAATACAGGAGTACCATTAACCTCTGGGTAGATGATTGTTATTTCCATTAACAGTCT 300
Qy 2044 ttgaacccaattcttatactctgacccaagacatttaagaagaatgattacatgcttc 2103
Db 301 TTGAACCCCAATTTCTATACCTGACCAACAGACATTTAAAGAAATGATTCATGCTG 360
Qy 2104 tggtaactacacagaagaataatctatgacacagaag--tcaagaataatgctc 2161
Db 361 TGGCAATTAATACAGCAAAAGAAATCTATAGACACAAAGATACGAAATATGCTC 420
Qy 2162 catcattcatctggttggaatggtgccaactgcaagagatgcaccctgattaa 2215
Db 421 CATCAATTCATCTGGGGGGAATGTGGCCACTCAGAGATGACCACTGATTA 474

RESULT 7
AX148176 1068 bp DNA linear PAT 08-JUN-2001
LOCUS AX148176 Sequence 17 from Patent WO0136471.
DEFINITION AX148176
ACCESSION AX148176
VERSION AX148176.1 GI:14347081
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 1068)
AUTHORS Chen,R., Dang,H.T. and Lowitz,K.P.
TITLE Endogenous and non-endogenous versions of human g protein-coupled
JOURNAL Patent: WO 0136471-A 17 25-MAY-2001;
Arena Pharmaceuticals, Inc. (US)
FEATURES
source location/Qualifiers
1..1068
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 278 a 215 c 214 g 361 t
ORIGIN

Query Match 12.2%; Score 437.2; DB 6; Length 1068;
Best Local Similarity 65.9%; Pred. No. 4.1e-79;
Matches 650; Conservative 0; Mismatches 333; Indels 3; Gaps 1;

Qy 1195 acgatggaattcatctctagagaatctcttgcaagacattatcaagagatttgc 1254
Db 10 ACGGACGGCATTTCTTCATTTGAGGACCTTGGCTAACAAATTCCTGCAAAATTTGTC 69
Qy 1255 tgggtgtatctgcaagttacatgcttggaaacattttgtcatctgcatgagacttat 1314
Db 70 TGGGTATATCTTTCATTTACCTGCTTGGAAATCTTTGTCATTTGGCATGAGATCTTTC 129
Qy 1315 atcagtcctgagaacaagctgtagcatgcaatattctctctgctgtgcccagctgc 1374
Db 130 ATTAAAGCTGAATAATACAACTGACGATGTCATTAATAATCCCTTGTGCGTGATGC 189
Qy 1375 ttaatggaatatattattctgcatcgtagggccttggaacaaagttctggtggaataac 1434
Db 190 CTGATGGGTATTACTTGTCTTGTGTCATTTTGCATTTTGCATTAATAATACGAGGAGAT 249
Qy 1435 aataagcatgcaagcctgtagatgtagagagtaactcatgttcaagctttagagcttggcc 1494
Db 250 CAGAGATATGCTTGTGATGAGAGCGGTGACAGCCGCTCATAGGGGTTCCCTGCC 309
Qy 1495 attctgtccacagaagatcagtttactgtttaacattctgcaattggaataataatcac 1554
Db 310 ATGCTGTCCACCGAAAGCTGTGTCCTACTGACCTACTTACTTGGGAAAGTTCTTG 369
Qy 1555 tgcatgtctatccctttagatgtagttagagactggaataatgcaagaataatcagttctg 1614
|||||


```

Db 370 GTCATGTCCTCCCTTCAGTAACATTGACCTGGAAAACGACACTCAGTACTCTC 429
      |||||  ||  ||  ||  |||||  ||  ||  ||
Qy 1615 attctcatttgatctactggtttatagtcgtcttcattcattgagcaataagaattt 1674
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 430 ATTTGCATCTGGATGGCGGGATTTTAAATACCTGTAATTCATTTTGGAAATAGGATTAT 489
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Qy 1675 ttcaaaactactatgagccaatgagatgatcttccttcattcagagaatagatacaga 1734
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 490 TTTCGAAACTTTTATGGGAAAAATGAGATGTTTCCACCTTTATTTATGACCAACAGAA 549
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Qy 1735 agtatggagccagattattcagtgagcaattttcttggtatttaatttgccgcaatt 1794
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 550 GATATTTGGAAACAAAGGATATCTCTTGGATTTTCCATGAGTGGAACCTCTGCGCTTT 609
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Qy 1795 atctcatagtttttctcataggaagcattgtttatagtgcttcataagagccataca 1854
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 610 CTCATCATGTTGTTTCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 669
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Qy 1855 gcaactgaataacgaatcaagtttaaaagaagatgatcccttgccaacggtttctt 1914
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 670 ACCACAGAAATAGGAATGTTTGGAGAGAGAGTGCTGTTCGAAATCGTTTCTTTT 729
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Qy 1915 atagattactatgacatctatgctgagataccattttgtagtgaatttcttcttactg 1974
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 730 ATAGTGTCTCTGATGCGCATCTGCTGATCTCTGATTTAGTAAATCCTTCCCTC 789
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Qy 1975 cttagagagaataacacaggtacataacctcttggtgtagtatttcttctccatt 2034
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 790 TTCCGGGGGAAATACCGACAGATGATCTCTGGATAGGATTTTCTTCCCTCAGTT 849
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Qy 2035 aacagtgttgaacccaattctctactactgaccacaagacattaaagaattt 2094
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 850 AACAGTGTGTAATCCATCTCTATCTCTCACACCACTTTTAAAGGACACAGTTG 909
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Qy 2095 catcggtttggtataacatacagaacaaatctatgacacagaacagatcagaacaa 2154
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 910 AAACAGTGTGACAAACA---TCAGAGAAATCAATTTTCAAAAATTAATAAAAAAAGT 966
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Qy 2155 tatgtccatcatcattcattcgtggtgga 2180
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 967 TTATCTACATCATCTGTTGTGATAGA 992
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```

REFERENCE AUTHORS TITLE JOURNAL COMMENT

RESULT 8
AC098607/c 171000 bp DNA linear HTG 20-DEC-2001
LOCUS Rattus norvegicus clone CH230-148B15, *** SEQUENCING IN PROGRESS
DEFINITION ***, 66 unordered pieces.
AC098607
AC098607.3 GI:17973780
VERSION HTG: HTGS, PHASE1.
KEYWORDS Norway rat.
SOURCE Rattus norvegicus
ORGANISM Eutheria; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 171000)
Munzy,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alstbrooks,S.L., Amaralunga,H.C., Are,J.R., Banks,T., Barbarella,J.,
Benton,J., Bimagne,K., Blankenburg,K., Bonnin,D., Boucek,J.,
Bowle,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burke,P., Burkett,C., Burrell,K.L., Byrd,N.C., Caron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,L., Chowdhury,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyler,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Day-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Eamhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escoto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
Hollins,B., Homsif,F., Howard,S., Huber,J., Hulik,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudan,S., Karlsson,E., Kelly,S., Khan,U., King,J., Kotvay,J.,
Kovar,C., Kratovic,J., Kueshi,A., Landry,N., Leal,B., Lewis,J.C.,
Lewis,L., Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,M.,
Loulsegad,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapa,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhinney,E., McLeod,M.P., Meador,M.,
Mel,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nokenko,S.,
Ogulu,M., Okunou,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoshitari,N.,
Slason,I., Sodergren,E., Sonike,T., Sparks,A., Stanley,H.,
Stone,H., Sulton,A., Swalek,A., Taber,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Watlington,S., Williams,G., Williamson,A., Wleczek,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G., and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 171000)
Worley,K.C.
Submitted (26-OCT-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Dec 20, 2001 this sequence version replaced gi:17064612.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GILC
Center clone name: CH230-148B15
----- Summary Statistics
Assembly program: Phrap; version 0.990329first call to
findPhrapList
Consensus quality: 139190 bases at least Q40
Consensus quality: 146928 bases at least Q30
Consensus quality: 152581 bases at least Q20
Estimated insert size: 137690; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 2.1x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank-draft.data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 66 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 7670: contig of 7670 bp in length
* 7671 7770: gap of unknown length
* 7771 12545: contig of 4775 bp in length
* 12546 12645: gap of unknown length
* 12646 16149: contig of 3504 bp in length
* 16150 16249: gap of unknown length
* 16250 22100: contig of 5851 bp in length
* 22101 22201: gap of unknown length
* 22201 26585: contig of 4385 bp in length
* 26586 26588: gap of unknown length
* 26588 32167: contig of 5482 bp in length
* 32167 32267: gap of unknown length

32268	37617:	contig of 5350	bp in length
37618	37717:	gap of unknown	length
37718	43062:	contig of 5345	bp in length
43063	43162:	gap of unknown	length
43163	46609:	contig of 3447	bp in length
46610	46709:	gap of unknown	length
46710	50626:	contig of 3917	bp in length
50627	50726:	gap of unknown	length
50727	54606:	contig of 3880	bp in length
54607	54706:	gap of unknown	length
54707	57421:	contig of 2715	bp in length
57422	57521:	gap of unknown	length
57522	60497:	contig of 2976	bp in length
60498	60597:	gap of unknown	length
60598	64460:	contig of 3863	bp in length
64461	64504:	gap of unknown	length
64561	66104:	contig of 3544	bp in length
68105	68204:	gap of unknown	length
68205	71451:	contig of 3247	bp in length
71452	71551:	gap of unknown	length
71552	74222:	contig of 2671	bp in length
74223	74322:	gap of unknown	length
74323	77237:	contig of 2915	bp in length
77238	77337:	gap of unknown	length
77338	79501:	contig of 2164	bp in length
79502	79601:	gap of unknown	length
79602	82022:	contig of 2421	bp in length
82023	82122:	gap of unknown	length
82123	84668:	contig of 2546	bp in length
84669	84768:	gap of unknown	length
84769	86880:	contig of 2112	bp in length
86881	86980:	gap of unknown	length
86981	88255:	contig of 2275	bp in length
88256	89365:	gap of unknown	length
89366	93173:	contig of 3818	bp in length
93174	93273:	gap of unknown	length
93274	96012:	contig of 2739	bp in length
96013	96112:	gap of unknown	length
96113	96812:	contig of 2500	bp in length
96813	98112:	gap of unknown	length
98113	101547:	contig of 2835	bp in length
101548	101647:	gap of unknown	length
101648	103910:	contig of 2263	bp in length
103911	104010:	gap of unknown	length
104011	107420:	contig of 3310	bp in length
107421	109694:	contig of 2274	bp in length
109695	109794:	gap of unknown	length
109795	111726:	contig of 1932	bp in length
111727	111826:	gap of unknown	length
111827	113814:	contig of 1988	bp in length
113815	113914:	gap of unknown	length
113915	115596:	contig of 1682	bp in length
115597	115696:	gap of unknown	length
115697	117489:	contig of 1793	bp in length
117490	117589:	gap of unknown	length
117590	119105:	contig of 1516	bp in length
119106	119205:	gap of unknown	length
119206	120898:	contig of 1693	bp in length
120899	120998:	gap of unknown	length
120999	123727:	contig of 2729	bp in length
123728	123827:	gap of unknown	length
123828	126119:	contig of 2292	bp in length
126120	126219:	gap of unknown	length
126220	127458:	contig of 1239	bp in length
127459	127558:	gap of unknown	length
127559	129450:	contig of 1892	bp in length
129451	129550:	gap of unknown	length
129551	133235:	contig of 2685	bp in length
132326	133335:	gap of unknown	length
132336	134307:	contig of 1972	bp in length
134308	135407:	gap of unknown	length
135409	135426:	contig of 1019	bp in length

Query Match	Best Local Similarity	Matches	6.9% 71.3%	Score 247.8; DB 2; Pred. No. 1,4e-40;	Length 171000;
			Conservative	0; Mismatches	152; Indels 21; Gaps
QY 1935	atcgtcgataccatctttagtgaatcttcctacgtctcagtagaataaccag	1994	135427	gap of unknown length	
DB 133291	AGGCGAGATATGCCCTTACAAATCGCACACATGCATCAATATGTTTCATCTTCTTCACGA	1332323	135522	gap of 226 bp in length	
QY 1995	taccataaccttgggtgaagtatttattcttcgccattaaagtgctttgaacccaat	2054	135522	contig of 226 bp in length	
DB 133231	TACTTTTACCTTGGGGGAGATTTTATTCCTCCATCAACAGTCTTTGAAGCCCAAT	1331727	137792	gap of unknown length	
QY 2055	tctctatactcgaccacaagaacatttaagaatgatcatcgtgcttggtaacta	2114	137793	gap of unknown length	
DB 133171	TATCTACACGTTTGACCACTTAACCTTTCAAGAAATGATTCATCAACTCTGGTCAACTA	1331121	137893	contig of 2232 bp in length	
QY 2115	cagacaaaagaatctatgacagcaaaag--tcagaaaaacatagctccatcatcat	2171	140125	gap of unknown length	
DB 133111	CAGACAAAGAGAGGCTGTGTGACAGAGAAAGGAGCTCAGAAACATACATCAATCATTCAT	1330552	140125	gap of unknown length	
QY 2172	ctgggtggaatgtggtgcacatgcagagagatgcacatcgagttaatgaagccgaccttt	2231	140225	contig of 1713 bp in length	
DB 133051	CTGGGTGGAATGTGGCCCTTGACAGGAGATGTCCACAGATCTTGAGAGCGGATGCTTT	1329922	141938	gap of unknown length	
QY 2232	cacataccctcggtgaatgtagtcaatgatttccatcaagaagatcaattccattcatg	2291	142038	gap of 1685 bp in length	
DB 132991	CACAGAACCCCTGTGATCTGTGCTTAGTTCCTCGGTCAATCTAGACTCAATTCCTTATTTATA	1329322	143723	gap of unknown length	
QY 2292	actgactctggaatcatcttcttcgacagagaatacttg999ggtctcattgaaggaatt	2351	143823	contig of 1293 bp in length	
DB 132931	ACTGGCTTTGGAAACCACTTATTTCTCA-AGAAATACTGT-GAAGTGTCTTAAAGGGGTTT	1328747	145215	gap of unknown length	
QY 2352	actggtatgaaatgaatatacacaanaatlaattatataatagctaaagaataatattt	2411	145216	gap of unknown length	
DB 132873	GCTGTATATG-AAGCAAAATACCAAAATATATCTTAAAGCAAGATTGCTTATATATGTGCA	1328155	146268	contig of 1553 bp in length	
QY 2412	acaagagacatgaggaataataaataatgactaatagtcttcaaaaggaatlaattatc	2471	146769	gap of unknown length	
DB 132814	CGAGG-----ACAGAAAGCTAGCTGTGCTTCATCAATCAAAAGAAAGCTAAATCAATCT	1327655	146869	gap of unknown length	
QY 2472	ataatagt-----ataatattagtagacatttgcatgaagaatgaaggaataatcact	2527	148780	contig of 1911 bp in length	
DB 132764	GACATATGTGTGCATGTATATATATAGTACATA-TTTCATACGACAGATGAAAGAAATCTATTT	1327066	150311	contig of 1432 bp in length	
			150312	gap of unknown length	
			150411	gap of unknown length	
			150412	contig of 1451 bp in length	
			151862	contig of 1451 bp in length	
			151863	gap of unknown length	
			151869	contig of 1907 bp in length	
			151963	contig of 1907 bp in length	
			153969	gap of unknown length	
			153970	contig of 1105 bp in length	
			155075	contig of 1105 bp in length	
			155174	gap of unknown length	
			155286	contig of 1112 bp in length	
			155175	contig of 1112 bp in length	
			156287	gap of unknown length	
			156387	gap of unknown length	
			157621	contig of 1235 bp in length	
			157721	gap of unknown length	


```

Db      2112 ACCTTAACAGCTTCAACATCAATGGAACAAAGATAGACTATTGGAAGAGACACATTC 2171
      893 caccctccagaacacgtagaattgattaggaagtaataagattgaacattccac 952
      2172 CACTTATGATACCTTAACAGTATTAGACCTGTAGACGAGGTGACCGCATGTGTATA 2231
      953 ccgttatatcaagagacgtgaagagcgtcaccaattgaactcttcctataatccatcc 1012
      2232 AAAACATGTTCAAAAGACCTGAAACAAATACACAGTTTAAACATATCCAGAAACACATCA 2291
      1013 agaaaattcaagcaaaccaattgattatctgtcaaaactcaagtcctcagcctagaag 1072
      2292 ATTCCATAGCAAAATGAGAGCAATTAATAATTATACCAATAGCCCTTGATGATTTGCTG 2351
      1073 ggaattaaattcaaatccaaacaaagagatgattagacctctatagaactctctcaca 1132
      2352 GCATATGATAAAGACATTTGGTCMAAAAGTGTTCATGGCTTACCAAGGCTTGTTGAGC 2411
      1133 tatatttaagaaattccagtaactgtggtatgacaccacagtctcgaactgttaaccaa 1192
      2412 TAAAACTGACACCTACCGGTTTGTGCTGCTCCAGAGGGTGTCAAGTGTCAACCCA 2471
      1193 acctgatgaattcatctctagagaatctcttgcaagcatatcagaagattttg 1252
      2472 AACCAATGATGTTTCCCTGCTGTGAAGACCTCATGTCCAAACCATGTGTGAGGGTCAGCA 2531
      1253 tctgggtgcatctgcagtaactgctgttggaacattttgcatltagatcgacct 1312
      2532 TCTGGGTGCTCGGTGTCATGCGACGTGCGGAACCTTTGTCGTCATCTTTGGAGGGTCA 2591
      1313 atatcagcttgagaacaaagcgtatgacacatgcaatcattctctcgtcgtgcagact 1372
      2592 GAGATTTTCAGAGAGAGAGAAAGTTCATCTTTTAATACCAACTTGGCCATGAGAGACT 2651
      1373 gcttaatggagaatatattatctgtagatggagagccttggaacaaagttcgtggagaat 1432
      2652 TTCTAATGGGGGCTACCTGTTTATCATCCACCCGCGACACTTACAGGGGGAGTGT 2711
      1433 acataagcagctgcgaagcgtgtgtagagagactacatgctgtgtagctttg 1492
      2712 ACATTCAGCCACGATGAGAACTGGAACAGATGAGACTCTGTGCTGCTGTTGCTGTCT 2771
      1493 ccaatctgcccagaagaatcagtttaactglttaacattctgacatitggaanaataca 1552
      2772 CCACATTTTCTAGTAATATATCAGTCTTGACTGTGTCACACATCACACTTGACCGCTCA 2831
      1553 tctgcatgtctatccttcttagatgtgtgagaccgtggaataatgcagaacaatacagttc 1612
      2832 TCTGCATCTCTGTTTCTGTGAGAGAACTGCACTTGCCCTGGAACAGGCCATCATTTGTCA 2891
      1613 tgaattcatattgattagctggtttatagtgcttcaatcattcattgagagaataaagaat 1672
      2892 TGAAGCTGTATATGGGTCCGAGTGTCTCTGCTGCTGCTTACCTCTGTTAAGATTTTCAT 2951
      1673 ttctcaaaaactactatgagcaaccaatgagatgcttccctctcattcagaagaatcacag 1732
      2952 ACTTGAGAACTCTTACGCGCGCTGTGAGATATGCTTGCGCCCTGCA---CGTAATCTCCAG 3008
      1733 aaagtgttggagccagattattatcagtggaatttttctgtgtatataatttggccgcat 1792
      3009 ACAGGGGGCCAGGCTGGAGTATCTGTGCTCTTCATCCCTTTCATCAATCTCTCTGTCA 3068
      1793 ttatcatcatagtttttctcattatgagaacatgtttatagtggtcacaagaatgacataa 1852
      3069 TCGTGTCTATAGCCTCTCTTACCTTTGATGATGTTTTCATAGCCAAAGACCCGAGAG 3128
      1853 cagcaactgaataacagaaatcaagttaaaaaagagaatgactcctgtccaaagcttttct 1912
      3129 CA-----GTGAGGACTGCTGAGAGCAAAATGACAAATGCAATGCGCCAGAGAAATGACT 3182
      1913 ttatagttattactgtagcatatgctgtgataacacatttctgtatgaaattcttccac 1972

```

```

Db      3183 TGATTGTGATGACTGACTTTTCTGCTGGGTGCCATCATTTGTCTTGCTTTTCTC 3242
      1973 tgcctcagtagaataaccaggtaccataaccctctggtagatatttatctgccca 2032
      3243 TACCGGTGCTAGGGCGGATGATCAGTGTATGATCATGATGATGCTGTGTTTGTCTCCCG 3302
      2033 ttacagctgcttgaaccaatctctatattctatccttgacaaagaacatt 2081
      3303 TGAACCTGCGACCAACCCGCTCATCTACACCTCTCCACAGCGCATTT 3351

RESULT 11
AL138708
LOCUS
DEFINITION
Human DNA sequence from clone RP11-432E15 on chromosome 13. Contains
STS, GSSs and a CpG island, complete sequence.
ACCESSION
AL138708.17 GI:11137618
VERSION
AL138708.17 GI:11137618
KEYWORDS
HTG; CpG island.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 170522)
REFERENCE
1 (bases 1 to 170522)
AUTHORS
Phillimore, B.
TITLE
Direct Submission
JOURNAL
Submitted (01-FEB-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquer@sanger.ac.uk
On Nov 10, 2000 this sequence version replaced gi:11121335.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WormPEP; Information
from the WormPEP database can be found at
http://www.sanger.ac.uk/projects/Celegans/WormPEP
This sequence was generated from part of bacterial clone contigs of human
chromosome 13, constructed by the Sanger Centre Chromosome 13
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr13
This sequence is the entire insert of clone RP11-432E15. The true
left end of clone RP11-154I23 is at 137612 in this sequence. The
true right end of clone RP11-95N14 is at 13246 in this sequence.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated repeat sequence elements. Where the sequence is
ambiguous, there is an annotation using the 'unsure' feature key.
RP11-432E15 is from the library RPCR-11.2 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBAC3.6.
FEATURES
source
1. 170522
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="13"
/clone="RP11-432E15"
/clone_1lb="RPCR-11.2"
19..289
/misc-feature
/note="match: GSS: Em:A0549338"
194..299
/repeat_region
/note="53 copies 2 mer tg 71% conserved"
208..299
/repeat_region
/note="23 copies 4 mer tgtg 77% conserved"
1456..1750
/repeat_region
/note="AlusX repeat: matches 3..300 of consensus"
2078..2183

```

```
misc_feature /note="MIR repeat: matches 6. .114 of consensus"
2600. .3122
/note="match: GSS: Em:AQ529216"
repeat_region /note="AluSq repeat: matches 1. .302 of consensus"
2646. .2940
/note="AluSq repeat: matches 1. .296 of consensus"
3590. .3894
/note="AluSq repeat: matches 1. .296 of consensus"
3926. .4462
/note="L2 repeat: matches 2233. .2750 of consensus"
4658. .4805
/note="L2 repeat: matches 2355. .2503 of consensus"
4911. .5224
/note="AluSq repeat: matches 1. .312 of consensus"
5237. .5332
/note="L2 repeat: matches 2575. .2694 of consensus"
5788. .6085
/note="AluJo repeat: matches 3. .301 of consensus"
7355. .7747
/note="sequence from overlapping clone bA15909
(A1136106).Assembly confirmed by restriction digest."
repeat_region /note="L2 repeat: matches 2 mer at 63% conserved"
7388. .7487
/note="50 copies 2 mer at 63% conserved"
7396. .7487
/note="23 copies 4 mer at 66% conserved"
7884. .7915
/note="MER20 repeat: matches 30. .62 of consensus"
7916. .8219
/note="AluJb repeat: matches 1. .302 of consensus"
8220. .8376
/note="MER20 repeat: matches 62. .210 of consensus"
8975. .8985
/note="sequence from overlapping clone
bA15909(A1136106).Assembly confirmed by restriction
digest. .9121
repeat_region /note="11 copies 4 mer caca 100% conserved"
9078. .9121
/note="complement(9603. .10104)
/note="match: GSS: Em:AQ569313"
11234. .11331
/note="sequence from overlapping clone bA15909
(A1136106).Assembly confirmed by restriction digest."
complement(11953. .12400)
/note="match: GSS: Em:AQ486928"
12477. .12601
/note="MIR repeat: matches 15. .138 of consensus"
12713. .12980
/note="sequence from overlapping clone bA15909
(A1136106).Assembly confirmed by restriction digest."
13272. .13417
/note="MIR repeat: matches 38. .188 of consensus"
complement(13670. .13919)
/note="match: GSS: Em:AQ021664"
complement(13740. .14130)
/note="match: GSS: Em:AQ032630"
13771. .14227
/note="sequence from overlapping clone bA15909
(A1136106).Assembly confirmed by restriction digest."
13864. .14087
/note="L2 repeat: matches 2515. .2750 of consensus"
16061. .16342
/note="MER1B repeat: matches 1. .244 of consensus"
16343. .16655
/note="AluSx repeat: matches 1. .312 of consensus"
16656. .16680
/note="MER1B repeat: matches 244. .305 of consensus"
complement(17019. .17275)
/note="match: GSS: Em:AQ452676"
17070. .17578
/note="match: GSS: Em:AQ511484"
17087. .17515
/note="match: GSS: Em:AQ0817416"
17278. .17794
/note="match: GSS: Em:AQ358157"

repeat_region 18301. .18381
/note="L1MB2 repeat: matches 6082. .6166 of consensus"
complement(18563. .19010)
misc_feature /note="match: GSS: Em:AQ242897"
20403. .20559
repeat_region /note="MER20 repeat: matches 1. .153 of consensus"
20870. .21069
/note="100 copies 2 mer tt 69% conserved"
20937. .21076
repeat_region /note="35 copies 4 mer ttcc 72% conserved"
21080. .21392
repeat_region /note="AluJb repeat: matches 1. .310 of consensus"
21395. .21538
repeat_region /note="MIR repeat: matches 94. .245 of consensus"
23716. .23960
/note="L2 repeat: matches 2381. .2617 of consensus"
23961. .24213
repeat_region /note="AluSx repeat: matches 47. .299 of consensus"
24214. .24342
repeat_region /note="L2 repeat: matches 2617. .2747 of consensus"
25636. .25785
repeat_region /note="MER5A repeat: matches 32. .189 of consensus"
25638. .25810
repeat_region /note="MER5A repeat: matches 17. .189 of consensus"
26013. .26227
repeat_region /note="MIR repeat: matches 10. .256 of consensus"
28426. .28484
/note="L2 repeat: matches 2641. .2700 of consensus"
28639. .28960
repeat_region /note="AluSx repeat: matches 1. .310 of consensus"
29689. .29861
/note="L2 repeat: matches 2006. .2186 of consensus"
complement(29720. .30187)
misc_feature /note="match: GSS: Em:AQ317477"
29901. .30075
repeat_region /note="MER5A repeat: matches 1. .184 of consensus"
complement(30119. .30414)
/note="match: GSS: Em:AQ230143"
32462. .32774
repeat_region /note="AluSx repeat: matches 1. .312 of consensus"
32807. .32877
/note="L2 repeat: matches 2635. .2707 of consensus"
33365. .33806
repeat_region /note="L2 repeat: matches 2248. .2742 of consensus"
complement(35852. .36413)
/note="match: GSS: Em:AQ284653
match: STS: Em:G53501"
complement(36043. .36433)
repeat_region /note="match: GSS: Em:AQ816559"
36420. .36469
/note="MIR repeat: matches 96. .148 of consensus"
36429. .36961
misc_feature /note="match: GSS: Em:AQ394779"
36452. .36845
/note="match: GSS: Em:AQ212353"
39203. .39242
repeat_region /note="20 copies 2 mer ca 100% conserved"
39250. .39648
repeat_region /note="L2 repeat: matches 2307. .2710 of consensus"
39703. .39750
repeat_region /note="24 copies 2 mer ca 75% conserved"
39859. .39976
/note="L1MC3 repeat: matches 7621. .7739 of consensus"
39988. .40283
repeat_region /note="AluSx repeat: matches 1. .296 of consensus"
40873. .41144
/note="AluJv repeat: matches 39. .310 of consensus"

Query Match 5.2%: Score 187.4; DB 9; Length 170522;
Best Local Similarity 64.8%; Pred. NO. 2.9e-28;
Matches 278; Conservative 0; Mismatches 151; Indels 0; Gaps 0;
QY 1349 tcattctctgtgtccgactgtctaagggatatattatctgtatcgagagct 1408
```



```

misc_feature      65270..75508
                  /note="assembly:fragment:00461"
misc_feature      75609..80749
                  /note="assembly:fragment:01454"
misc_feature      80850..101473
                  /note="assembly:fragment:00059"
misc_feature      101574..108695
                  /note="assembly:fragment:01172"
misc_feature      108796..120411
                  /note="assembly:fragment:01458"
misc_feature      120512..125324
                  /note="assembly:fragment:00440"
misc_feature      125425..129330
                  /note="assembly:fragment:00447"
misc_feature      129431..135864
                  /note="assembly:fragment:00564"
misc_feature      135965..156228
                  /note="assembly:fragment:00960"
misc_feature      156329..161861
                  /note="assembly:fragment:01066.0"
misc_feature      161962..164531
                  /note="assembly:fragment:01287"
misc_feature      164632..169371
                  /note="assembly:fragment:01319"
misc_feature      169472..176565
                  /note="assembly:fragment:01453"
BASE COUNT      52990 a 35385 c 33749 g 52527 t 1914 others
ORIGIN

```

```

Query Match      5.2%; Score 187.4; DB 2; Length 176565;
Best Local Similarity 64.8%; Pred. No. 2.9e-28;
Matches 278; Conservative 0; Mismatches 151; Indels 0; Gaps 0;

QY 1349 tcaattcctctgctgctgcgacgtccttaagggaataattatcgtgatcgaggact 1408
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 97765 TCAATTCCTCCACAGAGTCTGATTCCTGATGCGGCTTACTTCTTTGTCGCAATTT 97824

QY 1409 ttgaccctaagtttcgtgagagaataacaatgaatgcagctgtgagatgagagatctc 1468
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 97825 TCGATATATAAATACCGAGGCGAGTATCGAAGATATGCTTCTGCTGATGGAGACGTCG 97864

QY 1469 attgtcagctgtgagatccttgcgacatctgtccacagaagatcagtttactgttaa 1528
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 97885 AGTGGCGGCTCATGGGGTTCCTGGCGCATGCTGCCACGGAAGTCTGTTCTGCTACTGCA 97944

QY 1529 catttcgacattggaanaataacatcgtcattgtctatccctttagatgtgtgagacctg 1588
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 97945 CCTACTTCTGACTTTCGAGAGATTCCTCGTCAATGCTCTCCCTTCAGTAACATTCGACCTG 98004

QY 1589 gaaatgagagaacaattacagtttcgattcattcattgatttactgatttaagagctt 1648
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 98005 GAAAGCGGACAGCTCACTCATCTCCATTTGATCTGAGTGGCGGAGATTTTAAATAGCTG 98064

QY 1649 tcaattcctcattgagacaagaatttcaaaactactactgtgcacgaatgagatgct 1708
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 98065 TAATTCATTTTGGAAATAGAGTATTTTGGAAACTTTTATGGAAGAAATGAGATATGTT 98124

QY 1709 tcccttcattcagagatacagaagatlatgagccagatattatcagtgagcaattt 1768
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 98125 TCCCACTTATATGACCAACAGAGATATTTGGAAGCAAGAGGATTTCTCTTGAATTT 98184

QY 1769 ttcttggtta 1777
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 98185 TCCTAGGTA 98193

```

RESULT 13

```

AC098607          171000 bp   DNA       linear   HTG 20-DEC-2001
LOCUS             Rattus norvegicus clone CH230-148B15, *** SEQUENCING IN PROGRESS
DEFINITION        *** 66 unordered pieces.
ACCESSION         AC098607
VERSION           AC098607.3 GI:17973780
KEYWORDS          HTG; HTG; PHASE1.
SOURCE            Norway rats.
ORGANISM          Rattus norvegicus
                  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
                  Rattus.

```

REFERENCE

1 (bases 1 to 171000)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Albrooks,S.L., Amaratunga,H.C., Are,J.R., Banks,T., Barbata,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,D., Bowie,S., Briteva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Caron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homel,F., Howard,S., Huber,J., Huiyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louisaged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapa,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhinney,E., McLeod,M.P., Meador,M., Mel,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokewko,S., Oguh,M., Okwunou,G., Oragunye,N., Oyedero,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Ruiz,S., Ren,Y., Rives,M., Rojas,A., Rojiboken,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoshitari,N., Sisson,I., Sodergren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabot,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,Y., Villalón,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczky,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstein,G. and Gibbs,R.

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Unpublished
2 (bases 1 to 171000)
Direct Submission
Worley,K.C.
Submitted (26-OCT-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Dec 20, 2001 this sequence version replaced gi:17064612.

Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
Project Information
Center project name: GTCC
Center clone name: CH230-148B15
Summary Statistics
Assembly program: Phrap; version 0.990329first call to findphraplist
Consensus quality: 139190 bases at least Q40

Contact: saplens@watson.wustl.edu
 ----- Summary Statistics
 Center project name: H_NH0719M18

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
 Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:
 The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Moon, P.Y., Zhao, B., Frengen, E., Tatenos, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pleter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBAC3.6
 NEIGHBORING SEQUENCE INFORMATION:
 Actual start of this clone is at base position 1 of RP11-719M18; actual end is at base position 190122 of RP11-719M18.

FEATURES

```

source
    1..190122
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /chromosome="4"
    /map="4"
    /clone="RP11-719M18"
    /clone_lib="RPCI-11"
    505..553
    /rpt_family="AT_rich"
    repeat_region
    643..975
    /rpt_family="MaLR"
    repeat_region
    972..1001
    /rpt_family="(TG)n"
    repeat_region
    1015..1216
    /rpt_family="Alu"
    repeat_region
    1685..1853
    /rpt_family="MIR"
    repeat_region
    2241..2304
    /rpt_family="AT_rich"
    repeat_region
    2429..2833
    /rpt_family="L1"
    repeat_region
    3404..4027
    /rpt_family="ERV1"
    repeat_region
    4032..4422
    /rpt_family="ACHobo"
    repeat_region
    5753..5842
    /rpt_family="L1"
    repeat_region
    5843..6217
    /rpt_family="MaLR"
    repeat_region
    6225..6355
    /rpt_family="MaLR"
    repeat_region
    6353..6733
    /rpt_family="L1"
    repeat_region
    6734..7212
    /rpt_family="MBR1_type"
    repeat_region
    7213..7305
    /rpt_family="L1"
    repeat_region
    7809..8000
    /rpt_family="MIR"
    repeat_region
    8274..8748
    /rpt_family="MaLR"
    repeat_region
    8955..9081
    /rpt_family="L2"
    repeat_region
    9289..9579
    /rpt_family="Alu"
    repeat_region
    10261..10360
    /rpt_family="MaLR"
    repeat_region
    10446..10611
    /rpt_family="ERV1"
    repeat_region
    10614..11158
    /rpt_family="ERV1"
    repeat_region
    11280..11369
    /rpt_family="MBR1_type"
    repeat_region
    11400..11819
    /rpt_family="ERV1"
    repeat_region
    12071..12388
    /rpt_family="Alu"
    repeat_region
    12359..12405
    /rpt_family="(A)n"
    repeat_region
    12620..12973
    /rpt_family="MaLR"
    repeat_region
    13201..13221
    /rpt_family="AT_rich"
    repeat_region
    13207..13368
    /rpt_family="Alu"
    repeat_region
    13863..14381
    /rpt_family="ERV1"
    repeat_region
    15625..15659
    /rpt_family="AT_rich"
    repeat_region
    16105..16445
    /rpt_family="L2"
    repeat_region
    16584..16945
    /rpt_family="L1"
    repeat_region
    16934..17030
    /rpt_family="(CATATA)n"
    repeat_region
    19292..19314
    /rpt_family="(T)n"
    repeat_region
    19843..19954
    /rpt_family="T-rich"
    repeat_region
    24855..24934
    /rpt_family="L1"
    repeat_region
    24959..25593
    /rpt_family="ACHobo"
    repeat_region
    25594..25951
    /rpt_family="MaLR"
    repeat_region
    25952..26253
    /rpt_family="ACHobo"
    repeat_region
    26281..26515
    /rpt_family="L1"
    repeat_region
    26508..26530
    /rpt_family="AT_rich"
    repeat_region
    26935..26986
    /rpt_family="(TTTA)n"
    repeat_region
    26985..27435
    /rpt_family="L1"
    repeat_region
    27628..27653
    /rpt_family="AT_rich"
    repeat_region
    27761..27781
    /rpt_family="(TA)n"
    repeat_region
    27781..28057
    /rpt_family="Alu"
    repeat_region
    28062..28082
    /rpt_family="AT_rich"
    repeat_region
    28297..28358
    /rpt_family="CT-rich"
    repeat_region
    28322..28626
    /rpt_family="L1"
    repeat_region
    28741..28867

```

```
repeat_region /rpt_family="Alu"  
28875. .28931 /rpt_family="AT_rich"  
repeat_region 28930. .29909 /rpt_family="L1"  
repeat_region 29915. .30295 /rpt_family="L1"  
repeat_region 30351. .30585 /rpt_family="MALR"  
repeat_region 30864. .30953 /rpt_family="MIR"  
repeat_region 31260. .31675 /rpt_family="L1"  
repeat_region 31678. .32066 /rpt_family="L1"
```

Query Match 4.88; Score 173; DB 9; Length 190122;

Best Local Similarity 97.28; Pred. No. 2.5e-25;

Matches 176; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```
QY 1 ctgcttgaactgtaagattgagacagaaatagacacacacactgtagctgfatg 60  
|||  
Db 124029 CTGCTTGTAACTGCTAAGATTGACAGACAGAAATAGCACACACCACTGTGAGCTGATG 124088  
  
QY 61 cgattcagaacccaagaccaatttgcacttcaatcaatcagttgctcagatagaag 120  
|||  
Db 124089 CGATTGAGAAACCAAGACCAAAATTTGCTCCTTCATTAATCAGTGTGCTCAGATAGAG 124148  
  
QY 121 gaaatgacatcgtctctgtctcttctacacatcaatttttggaaaatatttctcat 180  
|||  
Db 124149 GAAATGACATCTGCTCTGCTCTCTTACATCTTAATTTTGGAAAATATTGTAGATAT 124208  
  
QY 181 g 181  
|  
Db 124209 G 124209
```

Search completed: September 5, 2002, 08:34:23
Job time: 20183 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 5, 2002, 03:56:10 : Search time 522.88 Seconds
(without alignments)
11768.324 Million cell updates/sec

Title: US-09-647-067-7
Perfect score: 3584
Sequence: 1 ctgcttgtaactgctaaga.....tagntttgtctgaaaccc 3584

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N.Geneseq_032802.*
1: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SID55/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SID55/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SID55/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SID55/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3580	99.9	3584	20	AA225346 Human LGR7 short f
2	2119.4	59.1	2467	20	AA225345 Human LGR7 long fo
3	1656.8	46.2	1804	21	AA290524 Human GPCR protein
4	1189.4	33.2	1191	22	AA006507 Human CON222 G pro
5	1000.4	27.9	1089	22	AA199557 Human expressed po
6	668.8	18.7	760	22	AA135687 Human musculoskele
7	580.4	16.2	636	21	AAA44932 Human secreted exp
8	528.6	14.7	530	22	AA199584 Human expressed po
9	528.6	14.7	530	22	AA135644 Human musculoskele

10	528.6	14.7	530	22	ABA06471 Human CDNA SEQ ID
11	528.6	14.7	530	22	AAS28950 CDNA encoding for
12	528.6	14.7	530	22	AAS29573 Human encoding po
13	528.6	14.7	530	22	AAS30187 DNA encoding rena
14	528.6	14.7	530	22	AAS34845 CDNA encoding nove
15	518.8	14.5	563	22	AA135575 Human musculoskele
16	439.2	12.3	474	22	AAF28059 Human GRL101 G-pr
17	437.2	12.2	1068	22	AA507943 Human CDNA encodin
18	382.8	10.7	420	21	AAC75735 Human OREX ORF1290
19	246.2	6.9	575	24	ABA05498 Human FSH-like GPC
20	220	6.1	494	20	AAV86595 EST clone BG43. H
21	181.6	5.1	429	24	AAS98139 Human DNA for pote
22	158.4	4.4	592	22	AAS46861 Human G protein-co
23	114.2	3.2	576	22	AAS46918 Human g protein-co
24	110	3.1	1018	22	AAH51001 Human nGPCR57 codi
25	96	2.7	1015	22	AAS57085 CDNA encoding Dros
26	96	2.7	1015	23	ABL07293 Drosophila melanog
27	76	2.1	1866	22	AAF29456 Human TANGO 325 cd
28	76	2.1	2114	22	AAS31666 Genomic sequence #
29	76	2.1	2114	22	AAS33455 DNA encoding human
30	76	2.1	2134	22	AAS21333 Human CDNA sequenc
31	76	2.1	2136	22	AAC82671 Human 7TM clone HD
32	76	2.1	2169	22	AAF29455 Human TANGO 325 cd
33	76	2.1	2180	22	AAS31607 CDNA encoding nove
34	76	2.1	2180	22	AAS33240 DNA encoding human
35	76	2.1	2204	22	AAD12983 Human neuronal gui
36	76	2.1	2516	22	AAS33136 DNA encoding human
37	76	2.1	2620	22	AAS31667 Genomic sequence #
38	76	2.1	2620	22	AAS33456 DNA encoding human
39	73.6	2.1	3174	23	ABL17583 Drosophila melanog
40	72.2	2.1	5616	23	ABL17582 Drosophila melanog
41	72.2	2.0	1080	22	AAS57089 CDNA encoding Dros
42	72.2	2.0	1080	23	ABL08799 Drosophila melanog
43	68.2	1.9	2019	22	AAH28471 Degenerate nucleot
44	63.8	1.8	1767	22	AAD05145 Human secreted pro
45	63.8	1.8	3578	22	AAD05121 Human secreted pro

ALIGNMENTS

RESULT 1
AA225346
ID AA225346 standard; CDNA; 3584 BP.
XX AA225346:
AC
XX
DT 20-DEC-1999 (first entry)
XX
DE Human LGR7 short form nucleotide sequence.
XX
KW Human; LGR4; LGR5; LGR7; G-protein coupled receptor; gene therapy;
KW extracellular leucine rich repeat region; mapping; identification; ss.
XX
OS Homo sapiens.
XX
PN W09948921-AL.
XX
PD 30-SEP-1999.
XX
PF 25-MAR-1999; 99WO-US06573.
XX
PR 26-MAR-1998; 98US-0079501.
XX
PA (STRD) UNIV LELAND STANFORD JUNIOR.
PA (ORGA) ORGANON NV.
PI Hsueh AJW, Hsu SY, Liang S, Van Der Spek PJ;
XX WPI: 1999-591074/50.
XX P-PDB: AAY42171.
XX
PT New G-protein coupled receptors, useful for identifying their own

PF ligands -
XX
XX Claim 4; Fig 4; 54pp; English.
XX

CC The present sequence encodes the human G-protein coupled receptor
CC having extracellular leucine rich repeat regions, designated LGR short
CC form. The LGR4, LGR5 and LGR7 proteins are used to identify ligands for
CC the receptor. The polypeptides and/or polynucleotides are also useful
CC for homologous or related genes, producing compositions that modulate
CC the expression or function of the receptors, gene therapy, mapping
CC functional regions of the receptors, studying associated physiological
CC pathways, in vivo prophylactic and therapeutic purposes, as immunogens
CC for producing antibodies, and for identifying biologically active
CC agents. The polypeptides contain a G-protein coupled seven
CC transmembrane region and a leucine rich repeat extracellular domain.
CC These regions capture and facilitate optimal orientation of its ligand.
CC The proteins are also expressed in diverse tissues.
XX
XX

SQ Sequence 3584 BP; 1124 A; 670 C; 647 G; 1139 T; 4 other;

Query Match 99.9%; Score 3580; DB 20; Length 3584;
Best Local Similarity 100.0%; Pred. No. 0;

Matches 3584; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctgcttgttaactgtcaagatgacagagaaatagacacacacactgtgagctgatatg 60
DB 1 ctgcttgttaactgtcaagatgacagagaaatagacacacacactgtgagctgatatg 60
QY 61 cgaatcagaacacagacacaaatttgcctcaacttcaatcaatgactgcagatagaag 120
DB 61 cgaatcagaacacagacacaaatttgcctcaacttcaatcaatgactgcagatagaag 120
QY 121 gaataaactctgtctgtctctctctcaacttaatttttgaaaataatttttcacat 180
DB 121 gaataaactctgtctgtctctctctcaacttaatttttgaaaataatttttcacat 180
QY 181 gggggtgagacagagatgtcaagtgctccctgtgctattcccccgttggaacatcaaaag 240
DB 181 gggggtgagacagagatgtcaagtgctccctgtgctattcccccgttggaacatcaaaag 240
QY 241 tgccttcctcagctcctctgcactgttaacggtgtgacgacgcggaatcggcgatgag 300
DB 241 tgccttcctcagctcctctgcactgttaacggtgtgacgacgcggaatcggcgatgag 300
QY 301 gaaactgt 360
DB 301 gaaactgt 360
QY 361 atgaaacatttacaagatgttccatcaggttcttccaatgttgagctgcaatgacttcag 420
DB 361 atgaaacatttacaagatgttccatcaggttcttccaatgttgagctgcaatgacttcag 420
QY 421 tggaaacttaataagaagcttccctctgtatgttccaagaatatacatatgtcttcagaag 480
DB 421 tggaaacttaataagaagcttccctctgtatgttccaagaatatacatatgtcttcagaag 480
QY 481 ctggactctgaaaaacataagattacatcatctcatctcatatgcttccagagactggaat 540
DB 481 ctggactctgaaaaacataagattacatcatctcatctcatatgcttccagagactggaat 540
QY 541 agccttactaactgtatctcagtcataacagaataactctctggaagccgggtgttttt 600
DB 541 agccttactaactgtatctcagtcataacagaataactctctggaagccgggtgttttt 600
QY 601 gaagatcttcacagactagaatgtgcgataattgaaagataatacactcagtcgaatttc 660
DB 601 gaagatcttcacagactagaatgtgcgataattgaaagataatacactcagtcgaatttc 660
QY 661 caacacaaattttatgactaataattctctatctctttagctctgtgtaataagctctc 720
DB 661 caacacaaattttatgactaataattctctatctctttagctctgtgtaataagctctc 720

QY 721 accggttaccgtgataacctctctgtcaacacatgcccagactacatgtgctgacct 780
DB 721 accggttaccgtgataacctctctgtcaacacatgcccagactacatgtgctgacct 780
QY 781 gaaggcaacatataccataatttaagaanaattgactttattcttcggcgatattaact 840
DB 781 gaaggcaacatataccataatttaagaanaattgactttattcttcggcgatattaact 840
QY 841 gttttgtatgaggaanaacaaataatactacattaaatgaanaatactttgcacctc 900
DB 841 gttttgtatgaggaanaacaaataatactacattaaatgaanaatactttgcacctc 900
QY 901 cagaacctgatatgaattgattagaagtaataagattgaaaatctccaccgttata 960
DB 901 cagaacctgatatgaattgattagaagtaataagattgaaaatctccaccgttata 960
QY 961 ttcaaggacttgaaggagctgtcacatgtgaatcttccataatccaatccagaanaatt 1020
DB 961 ttcaaggacttgaaggagctgtcacatgtgaatcttccataatccaatccagaanaatt 1020
QY 1021 caagcaaaccaatttattatcttgcactcaactcaagctctcagcttagaaggatgaa 1080
DB 1021 caagcaaaccaatttattatcttgcactcaactcaagctctcagcttagaaggatgaa 1080
QY 1081 attcaaatatccaaacaaaggatgttagaacctctatgaatcctctccacabataatt 1140
DB 1081 attcaaatatccaaacaaaggatgttagaacctctatgaatcctctccacabataatt 1140
QY 1141 aagaatttcagactgt 1200
DB 1141 aagaatttcagactgt 1200
QY 1201 ggaattcatctctagaagaatcctctgtgcaagcatattcagaagatttgcgtggt 1260
DB 1201 ggaattcatctctagaagaatcctctgtgcaagcatattcagaagatttgcgtggt 1260
QY 1261 gtaatcgcagttacactgtgtgtggaacatttgcatttgcatcgacctataacag 1320
DB 1261 gtaatcgcagttacactgtgtgtggaacatttgcatttgcatcgacctataacag 1320
QY 1321 tctgagaacaaagtgtatgtccatgtcaatcttctctgtcgtgtcgagcttcaatg 1380
DB 1321 tctgagaacaaagtgtatgtccatgtcaatcttctctgtcgtgtcgagcttcaatg 1380
QY 1381 ggaatatattatctgtgttcgagagcttgcgaactaaagtcttcgtgagaatacaataag 1440
DB 1381 ggaatatattatctgtgttcgagagcttgcgaactaaagtcttcgtgagaatacaataag 1440
QY 1441 catgacagctgtgtgagagagttactcaattgtcagctgttagagatcttggccattcg 1500
DB 1441 catgacagctgtgtgagagagttactcaattgtcagctgttagagatcttggccattcg 1500
QY 1501 tccaacagaagatcagtttactagttacaattctcgaatttggaanaataacatctgcat 1560
DB 1501 tccaacagaagatcagtttactagttacaattctcgaatttggaanaataacatctgcat 1560
QY 1561 gtctatccttttagatgtgtgagaccttggaanaatgagaacaaattacagttctgatttc 1620
DB 1561 gtctatccttttagatgtgtgagaccttggaanaatgagaacaaattacagttctgatttc 1620
QY 1621 atttggaattacgtggttttatagtgtcttcaatccatvtgagcaataagaaattttcaaa 1680
DB 1621 atttggaattacgtggttttatagtgtcttcaatccatvtgagcaataagaaattttcaaa 1680
QY 1681 aactactatggacacaaatgagatgtctccctcatcttcaggaanaatacagaagtatt 1740
DB 1681 aactactatggacacaaatgagatgtctccctcatcttcaggaanaatacagaagtatt 1740
QY 1741 ggaagccagatattatcagttggcaatttctctgtgattaaattggccgatttatcatc 1800
DB 1741 ggaagccagatattatcagttggcaatttctctgtgattaaattggccgatttatcatc 1800
QY 1801 atagtttttctcctatggaagcaltgttttatagtgtcatcaaaagtgcataacagcaact 1860

```
|||||
Db 1801 atagtttttccatggaagatgttttatagtgatcatcaaatgccaataacgcaact 1860
QY 1861 gaatacgggaatcaagtataaaagaatgaaccttgcgaagttcttcttatagta 1920
Db 1861 gaatacgggaatcaagtataaaagaatgaaccttgcgaagttcttcttatagta 1920
QY 1921 ttatcgtatgatatgtctgataccattttgttagtgaatttcttctcttcag 1980
Db 1921 ttatcgtatgatatgtctgataccattttgttagtgaatttcttctcttcag 1980
QY 1981 gtaaaaatacgaagttacataaacccttgggttagtgaatttcttctgcgaataag 2040
Db 1981 gtaaaaatacgaagttacataaacccttgggttagtgaatttcttctgcgaataag 2040
QY 2041 gctttgaacccaattctctatactctgacacaagaaccatttaagaatgatactcgg 2100
Db 2041 gctttgaacccaattctctatactctgacacaagaaccatttaagaatgatactcgg 2100
QY 2101 ttttggfataactacagacaagaatactatgagacagcaaggctcagaataactatgct 2160
Db 2101 ttttggfataactacagacaagaatactatgagacagcaaggctcagaataactatgct 2160
QY 2161 ccaatcatcatctggtggtgaatgtggccaactgaggaagatgccaactgagttatgag 2220
Db 2161 ccaatcatcatctggtggtgaatgtggccaactgaggaagatgccaactgagttatgag 2220
QY 2221 ccggaacttttccataaccctgtgaaatgtcactgatttctcaatacagaagactcaat 2280
Db 2221 ccggaacttttccataaccctgtgaaatgtcactgatttctcaatacagaagactcaat 2280
QY 2281 tccattcatgacgactcgtaaattcaattcttcgcagagaatactgtgggggtgcttc 2340
Db 2281 tccattcatgacgactcgtaaattcaattcttcgcagagaatactgtgggggtgcttc 2340
QY 2341 atgagggatttactgtgatgaaaaatgaatacacaaaaatattatataataatagctag 2400
Db 2341 atgagggatttactgtgatgaaaaatgaatacacaaaaatattatataataatagctag 2400
QY 2401 ataaatttttacaagaacatgagaagaataaaataaaatgactaaagctcttcaaggaa 2460
Db 2401 ataaatttttacaagaacatgagaagaataaaataaaatgactaaagctcttcaaggaa 2460
QY 2461 gtaattatatacaatgaatgatataatagtagagacatttgcataagaataatagaaga 2520
Db 2461 gtaattatatacaatgaatgatataatagtagagacatttgcataagaataatagaaga 2520
QY 2521 tctacttcgaatacatcatcatcttcttcaacatgcatlcttatgtagaccactactat 2580
Db 2521 tctacttcgaatacatcatcatcttcttcaacatgcatlcttatgtagaccactactat 2580
QY 2581 gtgcatagcattgcataatagctctggaatgagacagtgcagaaccccttcaactgtaga 2640
Db 2581 gtgcatagcattgcataatagctctggaatgagacagtgcagaaccccttcaactgtaga 2640
QY 2641 taggttttaagacaagaactatatacaaatgccttgcacttccctagtttaagaatag 2700
Db 2641 taggttttaagacaagaactatatacaaatgccttgcacttccctagtttaagaatag 2700
QY 2701 cttaacctgtcatgtgcatcagcaagaatcagacatttcaataaaggtttaaaagt 2760
Db 2701 cttaacctgtcatgtgcatcagcaagaatcagacatttcaataaaggtttaaaagt 2760
QY 2761 ttggaatactcagtgatattgcatacatagaaaatgctcagctgtttgcaaaaataatc 2820
Db 2761 ttggaatactcagtgatattgcatacatagaaaatgctcagctgtttgcaaaaataatc 2820
QY 2821 tgtttaagaatcatcttactctctttaaagttccatacaacttgaagcacaacacac 2880
Db 2821 tgtttaagaatcatcttactctctttaaagttccatacaacttgaagcacaacacac 2880
QY 2881 atatttataaagaatgcttgcatagaacccaagaacacttcttgcacact 2940
Db 2881 atatttataaagaatgcttgcatagaacccaagaacacttcttgcacact 2940
```

```
Db 2881 atatttataaagaatgcttgcatagaacccaagaacacttcttgcacact 2940
QY 2941 tctgtcccaagtttctcttgcatttaaatgaatacatcatatggaatttggaaatagaga 3000
Db 2941 tctgtcccaagtttctcttgcatttaaatgaatacatcatatggaatttggaaatagaga 3000
QY 3001 gtatgagtacgagagaagtgagatcagaanaaacatgaatgaatgaataacattacatta 3060
Db 3001 gtatgagtacgagagaagtgagatcagaanaaacatgaatgaatgaataacattacatta 3060
QY 3061 gtggaacacctcctgaataaataccttgatgtgacgttaactgatttccaagaagatgcc 3120
Db 3061 gtggaacacctcctgaataaataccttgatgtgacgttaactgatttccaagaagatgcc 3120
QY 3121 aagacaaaaggctttccaanaacccgtgctgttttaagaacagactaagtgtttaat 3180
Db 3121 aagacaaaaggctttccaanaacccgtgctgttttaagaacagactaagtgtttaat 3180
QY 3181 tcaccacattagaatgggtgagatgttattggtgtgtaatactccagttaaagcagttcaa 3240
Db 3181 tcaccacattagaatgggtgagatgttattggtgtgtaatactccagttaaagcagttcaa 3240
QY 3241 aggaanaaagcgtggaatgacatgattcagaacttaattcagaaggaagaaagctgta 3300
Db 3241 aggaanaaagcgtggaatgacatgattcagaacttaattcagaaggaagaaagctgta 3300
QY 3301 tgtacacatttcaatttaagaagaatacttcttcaagaataatgacttaattctcttt 3360
Db 3301 tgtacacatttcaatttaagaagaatacttcttcaagaataatgacttaattctcttt 3360
QY 3361 gcactgcagacgtgagatcactaacttttaactagttgtcttctcactacactcgt 3420
Db 3361 gcactgcagacgtgagatcactaacttttaactagttgtcttctcactacactcgt 3420
QY 3421 tattagnatttttgccttcaataatgtaaacctttaaagcagagaagaataatgttttca 3480
Db 3421 tattagnatttttgccttcaataatgtaaacctttaaagcagagaagaataatgttttca 3480
QY 3481 gatagtttcaaatnccnaaaatggttgcagacaaaataactcggatcnaaccataat 3540
Db 3481 gatagtttcaaatnccnaaaatggttgcagacaaaataactcggatcnaaccataat 3540
QY 3541 gcccttattgaatatatagttgataagnttgttctgtaaaacc 3584
Db 3541 gcccttattgaatatatagttgataagnttgttctgtaaaacc 3584

RESULT 2
AA25345
ID AA25345 standard; cDNA; 2467 BP.
XX
AC AA25345;
XX
DT 20-DEC-1999 (first entry)
XX
DE Human LGR7 long form nucleotide sequence.
XX
KW Human; LGR4; LGR5; LGR7; G-protein coupled receptor; gene therapy;
KW extracellular leucine rich repeat region; mapping; identification; ss.
OS Homo sapiens.
XX
PN W09948921-A1.
XX
PD 30-SEP-1999.
XX
PE 25-MAR-1999; 99WO-US06573.
XX
PR 26-MAR-1998; 98US-0079501.
XX
PA (STRD ) UNIV LEELAND STANFORD JUNIOR.
PA (ORGA ) ORGANON NV.
XX
```



```
OY 1636 ttatagtggttcatccatcgtgagcaataagaattttcaaaactactatgcaac 1695
    |||||
DB 1811 ttatagtggttcatccatcgtgagcaataagaattttcaaaactactatgcaac 1870
OY 1666 aatgagatgcttcctccttcattcagaagatacagaagatgtgagccagattat 1755
    |||||
DB 1871 aatgagatgcttcctccttcattcagaagatacagaagatgtgagccagattat 1930
OY 1766 tcagtggaattttctctgtgattatttggccgatttatcatcatggtttttcccat 1815
    |||||
DB 1931 tcagtggaattttctctgtgattatttggccgatttatcatcatggtttttcccat 1990
OY 1816 ggaacacgtttatagtggttcatcaagaatgccaacagacgaataacagatcaaa 1875
    |||||
DB 1991 ggaacacgtttatagtggttcatcaagaatgccaacagacgaataacagatcaaa 2050
OY 1876 gtaaaaaagagatgcttcctgccaacggtttttcttattagtaattactatgata 1935
    |||||
DB 2051 gtaaaaaagagatgcttcctgccaacggtttttcttattagtaattactatgata 2110
OY 1936 tgcgtgataccattttgttagtgaatttcttcactgctcaggtagaataaccaggt 1995
    |||||
DB 2111 tgcgtgataccattttgttagtgaatttcttcactgctcaggtagaataaccaggt 2170
OY 1996 accataacctgtgtagtgaatttttcttgcaccttaacagagctttagaaccgaatt 2055
    |||||
DB 2171 accataacctgtgtagtgaatttttcttgcaccttaacagagctttagaaccgaatt 2230
OY 2056 ctctactctgaccacagaaccatttaagaatgattcactcgttttggatatactac 2115
    |||||
DB 2231 ctctactctgaccacagaaccatttaagaatgattcactcgttttggatatactac 2290
OY 2116 agacaagaataatctatgagcagcaaggtcagaacacatattgctcatcattctgg 2175
    |||||
DB 2291 agacaagaataatctatgagcagcaaggtcagaacacatattgctcatcattctgg 2350
OY 2176 gtggaagtgtgcccactgcaagagatgcccactgagtaatgaagccgacctttcaca 2235
    |||||
DB 2351 gtggaagtgtgcccactgcaagagatgcccactgagtaatgaagccgacctttcaca 2410
OY 2236 tacacctgtgaatgtcactgattctcaatcaacgagactaatctccatcatga 2292
    |||||
DB 2411 tacacctgtgaatgtcactgattctcaatcaacgagactaatctccatcatga 2467

RESULT 3
AAZ90524
ID AAZ90524 standard; cDNA: 1804 BP.
XX
AC AAZ90524;
XX
DT 05-JUN-2000 (first entry)
XX
DE Human GPCR protein (HGRP) encoding cDNA (clone ID 2488822).
XX
KW Human; G protein coupled protein receptor; HGRP; cell proliferation;
    neurologic; immune disorder; cytosolic; anti-arteriosclerotic;
    anti-atherosclerotic; hepatotropic; antiinflammatory; virucide; leukemia;
    immunomodulatory; anemia; asthma; gastrointestinal; anti-epileptic;
    anti-Alzheimer's; anti-Parkinsonian; gene therapy; ss.
XX
OS Homo sapiens.
PN WO200015793-A2.
PD 23-MAR-2000.
XX
PE 17-SEP-1999; 99WO-US20958.
PR 17-SEP-1998; 98US-0156513.
PA (INCY-) INCYTE PHARM INC.

XX
PI Bandman O, Lal P, Tang YT, Corley NC, Guegler KJ, Gorgone GA;
    Baughn MR;
XX
DR WPI: 2000-271432/23.
XX
P-PSDB: AAY57286.
XX
PT Human G protein coupled protein receptor peptides useful for the
    prevention, diagnosis and treatment of cell proliferative, neurological
    and immune disorders -
XX
PS Claim 9; Page 69-70; 71pp; English.
XX
CC The invention provides human G protein coupled protein receptor (HGRP)
    polypeptides and polynucleotides encoding them. The polypeptides can be
    produced by standard recombinant methodology. The polynucleotides and
    polypeptides may be used in the prevention, treatment and diagnosis of
    diseases associated with their inappropriate expression. Diseases that
    can be treated are cell proliferative disorders (e.g. arteriosclerosis,
    atherosclerosis and hepatitis), cancers (e.g. leukemia, melanomas and
    adenocarcinoma), immune disorders (e.g. anemia, asthma and Crohn's
    disease) and neurological disorders (e.g. epilepsy, Alzheimer's disease
    and Parkinson's disease). The anti-HGRP antibodies may also be used as
    diagnostic agents for detecting the presence of HGRP polypeptides in
    samples (e.g. by enzyme linked immunosorbent assay (ELISA)). Sequences
    AAZ90521-526 represent cDNA fragments encoding the HGRP polypeptides
    (AAZ57283-288).
XX
SQ Sequence 1804 BP; 563 A; 318 C; 327 G; 596 T; 0 other;

Query Match 46.2%; Score 1656.8; DB 21; Length 1804;
Best Local Similarity 99.2%; Pred. No. 0;
Matches 1675; Conservative 0; Mismatches 12; Indels 1; Gaps 1;

OY 1057 tctccagcctgaagggaattgaattcaaatccacaagaagatgttagacctt 1116
    |||||
DB 105 ttaacagcctgaagggaattgaattcaaatccacaagaagatgttagacctt 164
OY 1117 atgaatctctcacatatattttaagaatccagtagtggatgacacacatgtt 1176
    |||||
DB 165 ggaatctctcacatatattttaagaatccagtagtggatgacacacatgtt 224
OY 1177 cgcacgtgaaacccaacacgtatgaattcatctctagaatctcttggcaagcat 1236
    |||||
DB 225 cgcacgtgaaacccaacacgtatgaattcatctctagaatctcttggcaagcat 284
OY 1237 attcagagatattgtctgggtgtatctgcagttacacctgttggaaacattttgtc 1296
    |||||
DB 285 attcagagatattgtctgggtgtatctgcagttacacctgttggaaacattttgtc 344
OY 1297 attgcatgagacattatatacagtcgtgaaacagctglatgcatatcatctt 1356
    |||||
DB 345 attgcatgagacattatatacagtcgtgaaacagctglatgcatatcatctt 404
OY 1357 cctcgtgtgcccactgtctaatgggaatatattatctgtatogggagctttgacct 1416
    |||||
DB 405 cctcgtgtgcccactgtctaatgggaatatattatctgtatogggagctttgacct 464
OY 1417 aagttcgtgggaatacacaataagcagtcgcagctgtgtgtagagatcctcatgtcag 1476
    |||||
DB 465 aagttcgtgggaatacacaataagcagtcgcagctgtgtgtagagatcctcatgtcag 524
OY 1477 ctgttagaactcttgccactctgtccacagaagatacagtttactgttaacattctg 1536
    |||||
DB 525 ctgttagaactcttgccactctgtccacagaagatacagtttactgttaacattctg 584
OY 1537 acattggaataatacatcgtcgtgtctatccttttagtgtgtgagactgtgaaatgc 1596
    |||||
DB 585 acattggaataatacatcgtcgtgtctatccttttagtgtgtgagactgtgaaatgc 644
OY 1597 agaacatacagtcctgattcattctgattactgttattatagtgcttcatca 1656
    |||||
```

Db 645 agaacaattacagttctgattctcattcttgatgtatttagttagtcttcacaa 704
 Oy 1657 ttgagcaataaggaatttttaaaactatgtgcacaaatgagtagtctccctctt 1716
 Db 705 ttgagcaataaggaatttttaaaactatgtgcacaaatgagtagtctccctctt 764
 Oy 1717 cattcagaagatagcagaagatttggagccagattatctcaatgtgcattttctt 1776
 Db 765 cattcagaagatagcagaagatttggagccagattatctcaatgtgcattttctt 824
 Oy 1777 attaatgtgcgcgcatattatcatctatgtttttccatggagaatgtttatagtt 1836
 Db 825 attaatgtgcgcgcatattatcatctatgtttttccatggagaatgtttatagtt 884
 Oy 1837 catcaaatgtccataacgcaactgaactgaatacagaatcaagttaaaaagagatgcctt 1896
 Db 885 catcaaatgtccataacgcaactgaactgaatacagaatcaagttaaaaagagatgcctt 944
 Oy 1887 gccaaaagttttctttatagttatctatgtatctgatactatgctggataccattttgta 1956
 Db 945 gccaaaagttttctttatagttatctatgtatctgatactatgctggataccattttgta 1004
 Oy 1957 gtgaaattcttctcactccttcaggttagaataacaggttacatgactcttggtagtg 2016
 Db 1005 gtgaaattcttctcactccttcaggttagaataacaggttacatgactcttggtagtg 1064
 Oy 2017 attttatctgcgcattacagtgcttgaacccaattctctatctctgacccaaga 2076
 Db 1065 attttatctgcgcattacagtgcttgaacccaattctctatctctgacccaaga 1124
 Oy 2077 ccatttaagaagaatgattcatcggttttgatatacctaagaagaataatctatggac 2136
 Db 1125 ccatttaagaagaatgattcatcggttttgatatacctaagaagaataatctatggac 1184
 Oy 2137 agcaaaagtcagaaacataatgctccatcatctctggtggaatgtggccatgtgag 2196
 Db 1185 agcaaaagtcagaaacataatgctccatcatctctggtggaatgtggccatgtgag 1244
 Oy 2197 gagatgccaactgtgtaatgaagccggaactttcacatacccttggaatgtgacatg 2256
 Db 1245 gagatgccaactgtgtaatgaagccggaactttcacatacccttggaatgtgacatg 1304
 Oy 2257 atttctcaatacagaagactcaattcctcatgaatgactgtaactcgaattcattcttcg 2316
 Db 1305 atttctcaatacagaagactcaattcctcatgaatgactgtaactcgaattcattcttcg 1364
 Oy 2317 cagaagaactgctgggggtgcttcaatggaggttactgctgcatggaatgataccacaa 2376
 Db 1365 cagaagaactgctgggggtgcttcaatggaggttactgctgcatggaatgataccacaa 1423
 Oy 2377 aattaattataataatagctaagaataaattttacagaagacatgaggaataaataaa 2436
 Db 1424 aattaattataataatagctaagaataaattttacagaagacatgaggaataaataaa 1483
 Oy 2437 tgactaatgctccttaacaaaggaagtaattatatacaataatgtatatattagtagaca 2496
 Db 1484 tgactaatgctccttaacaaaggaagtaattatatacaataatgtatatattagtagaca 1543
 Oy 2497 ttgtgcataagaattgaagaataatctactcagtaaatcatcatttttctaactg 2556
 Db 1544 ttgtgcataagaattgaagaataatctactcagtaaatcatcatttttctaactg 1603
 Oy 2557 cattattgaatccacatactatgtgataagcatgtgcaatatagtcctgtggaatgaca 2616
 Db 1604 cattattgaatccacatactatgtgataagcatgtgcaatatagtcctgtggaatgaca 1663
 Oy 2617 gtgcagaaccttcaatctgtagatggtttaatgacaaaagactatacaaaatgcattc 2676
 Db 1664 gtgcagaaccttcaatctgtagatggtttaatgacaaaagactatacaaaatgcattc 1723
 Oy 2677 tgcagttcctagtttaagtagagcttactcgtcatgctgcatagcagaagataatagc 2736
 Db 1724 tgcagttcctagtttaagtagagcttactcgtcatgctgcatagcagaagataatagc 1783

Oy 2737 atttttaa 2744
 Db 1784 atttttaa 1791
 RESULT 4
 AAD06507
 ID AAD06507 standard; DNA; 1191 BP.
 AC AAD06507;
 DT 10-AUG-2001 (first entry)
 XX
 DE Human CON222 G protein-coupled receptor DNA.
 KW Human; G protein-coupled receptor; GPCR; CON222 protein; schizophrenia; neuroleptic; neurotropic; neuroprotective; bipolar disease; psychotropic; neurological disorder; psychiatric disease; neurosis; anxiety; neuritis; attention deficit hyperactivity disorder; neurosthenia; senile dementia; affective disorder; neuropathy; Alzheimer's disease; Parkinson's disease; depression; migraine; genetic screening; ds.
 KW Homo sapiens.
 OS
 XX
 FH Key Location/Qualifiers
 FT CDS 1..1191
 FT /*tag= a
 FT /product= "Human G protein-coupled receptor protein"
 XX
 PN W0200131014-A2.
 XX
 PD 03-MAY-2001.
 XX
 PR 27-OCT-2000; 2000WO-US29601.
 XX
 PR 27-OCT-1999; 99US-0427653.
 PR 27-OCT-1999; 99US-0427859.
 PR 27-OCT-1999; 99US-0428020.
 PR 27-OCT-1999; 99US-0428114.
 PR 28-OCT-1999; 99US-0429517.
 PR 28-OCT-1999; 99US-0429555.
 PR 28-OCT-1999; 99US-0429676.
 PR 28-OCT-1999; 99US-0429695.
 PR 03-DEC-1999; 99US-0454399.
 PR 12-JAN-2000; 2000US-0481794.
 XX
 PA (PHAA) PHARMACIA & UPJOHN CO.
 XX
 PI Vogell G, Wood LS, Merchant K;
 XX
 DR WPI: 2001-32853/34.
 DR P-PSDB: AAE02498.
 XX
 PT Seven transmembrane receptor polypeptides and polynucleotides, useful for treating neurological or psychiatric disorders, e.g. schizophrenia, as well as for identifying compounds useful for treating schizophrenia
 PT
 PS Claim 22; Page 15-16; 215pp; English.
 XX
 CC The invention relates to human G protein-coupled receptor (GPCR) and their corresponding DNA molecules. GPCR is also referred as seven transmembrane receptor. G protein-coupled receptor protein is useful for treating neurological disorder, particularly schizophrenia. GPCR protein is also useful for identifying compounds useful for treating schizophrenia. These compounds are also useful for treating other neurological and psychiatric diseases, e.g. depression, anxiety, bipolar disease, affective disorders, attention deficit hyperactivity disorder/attention deficit disorder, epilepsy, neuritis, neurosthenia, neuropathy, CC neurosis, Alzheimer's disease, Parkinson's disease, migraine and senile dementia. The invention also provides genetic screening procedures that entail analysing a person's genome with respect to GPCR. The vectors are

CC useful for the recombinant production of the GPCR's. The present DNA
CC sequence encodes human CON222 G protein-coupled receptor (GPCR) protein.
XX
SQ Sequence 1191 BP; 340 A; 229 C; 226 G; 396 T; 0 other;

Query Match 33.2%; Score 1189.4; DB 22; Length 1191;
Best Local Similarity 99.9%; Pred. No. 6,9e-268;
Matches 1190; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
QY 1102 atgttagacctctatgaaatctctcaacatattttaagaattccagctatgag 1161
D 1 atgttagacctctggaatctctcaacatattttaagaattccagctatgag 60
QY 1162 tatgaccacatgtctgcaagctgtgaaacacacatgtaatttcaatctctagagaat 1221
D 61 tatgaccacatgtctgcaagctgtgaaacacacatgtaatttcaatctctagagaat 120
QY 1222 ctcttgcaagcatctatccagagagatttgctgggttgatctgcaagctatgctt 1281
D 121 ctcttgcaagcatctatccagagagatttgctgggttgatctgcaagctatgctt 180
QY 1282 ggaacacatttctgcatctgcatgcaacctatatacagctgagaaacagctgtagcc 1341
D 181 ggaacacatttctgcatctgcatgcaacctatatacagctgagaaacagctgtagcc 240
QY 1342 atgtgcaatctctctctctgctgctgcaagctgtaattgggaatatatttctgtgac 1401
D 241 atgtgcaatctctctctctgctgctgcaagctgtaattgggaatatatttctgtgac 300
QY 1402 ggaagcttgacacaaagcttctgctgagaaatacaataagcctgcaagctgtgtagag 1461
D 301 ggaagcttgacacaaagcttctgctgagaaatacaataagcctgcaagctgtgtagag 360
QY 1462 agtactaatgtcagctgtgtagaactcttgccatctctgccaagaagtagtcgtttaa 1521
D 361 agtactaatgtcagctgtgtagaactcttgccatctctgccaagaagtagtcgtttaa 420
QY 1522 ctgttaacatttctgcaattggaaataacatctgcaattgtctactctttagatgtgtg 1581
D 421 ctgttaacatttctgcaattggaaataacatctgcaattgtctactctttagatgtgtg 480
QY 1582 agacctggaataatgcagaaacatctacagtttctgatttccatttgatctggtttaa 1641
D 481 agacctggaataatgcagaaacatctacagtttctgatttccatttgatctggtttaa 540
QY 1642 gtggcttcatctccatgtgagcaataaagaaatttccaaaaactactatgaccaaagtag 1701
D 541 gtggcttcatctccatgtgagcaataaagaaatttccaaaaactactatgaccaaagtag 600
QY 1702 gtatgcttccctctcttcttcaagaagatacagaagaatttggagcccgattatctcagt 1761
D 601 gtatgcttccctctcttcttcaagaagatacagaagaatttggagcccgattatctcagt 660
QY 1762 gcaatttctgtgtatatttgccgcatcttcatcatagtttcttccatggaagc 1821
D 661 gcaatttctgtgtatatttgccgcatcttcatcatagtttcttccatggaagc 720
QY 1822 atgtttcatagtgctcatcaaatgtccataacagcaactgtaatacagaaatcaaa 1881
D 721 atgtttcatagtgctcatcaaatgtccataacagcaactgtaatacagaaatcaaa 780
QY 1882 aaagaagatgactctgccaacggttttctttagtatttactgtagctatgctgg 1941
D 781 aaagaagatgactctgccaacggttttctttagtatttactgtagctatgctgg 840
QY 1942 ataccacatttctgtagaataattcttcaactgctcaggtagagaataccagtagcaca 2001
D 841 ataccacatttctgtagaataattcttcaactgctcaggtagagaataccagtagcaca 900
QY 2002 acctctgggtagtagattttattctctcccaataacagtgcttgaaccacaattctctat 2061
D 901 acctctgggtagtagattttattctctcccaataacagtgcttgaaccacaattctctat 960
```

```
QY 2062 acctgaccacaagaccatttaagaatagtatcatcggttggtagtaactacagaca 2121
D 961 acctgaccacaagaccatttaagaatagtatcatcggttggtagtaactacagaca 1020
QY 2122 agaaatctatgtagcagcaaaagctcagaaaacatactccatcatcatcctggtgaa 2181
D 1021 agaaatctatgtagcagcaaaagctcagaaaacatactccatcatcatcctggtgaa 1080
QY 2182 atgtggccactgtagagagatgccacctgagtttaatgaagccggaacctttacataccc 2241
D 1081 atgtggccactgtagagagatgccacctgagtttaatgaagccggaacctttacataccc 1140
QY 2242 tgtaaatgtacatgatttccatcaacagagactcaatttcctattga 2292
D 1141 tgtaaatgtacatgatttccatcaacagagactcaatttcctattga 1191

RESULT 5
AA199557
ID AA199557 standard; cDNA; 1089 BP.
XX
AC AA199557;
XX
DT 04-JAN-2002 (first entry)
XX
DE Human expressed polynucleotide SEQ ID NO 20.
XX
KW Human; nootropic; neuroprotective; cytosolic; dermatological; virocid; immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary; antiparkinsonian; antischling; antianemic; antiallergic; cancer; antirheumatic; hepatotropic; cerebroprotective; antiinflammatory; antiallergic; antidiabetic; antilicer; anticonvulsant; antifungal; antiparasitic; cardiac; immune disorder; cardiovascular disorder; neurological disease; infection; nephrotropic; gene therapy; vaccine; ss.
XX
OS Homo sapiens.
XX
PN W020015387-A1.
XX
PD 02-AUG-2001.
XX
PE 17-JAN-2001; 2001WO-US01310.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
```



```

Db 181 attgattattctgtcaaccctcaagctctct---cctagaaggagattgaattccaatat 237
Oy 1092 ccaacaagaagatggttagaacctctatgaatcctctctacacataatttaagaatacca 1151
Db 238 ccaataaagaagtgttagaacctctatgaatcctctctacacataatttaagaatacca 297
Oy 1152 gtagctggtggaatgacacacatgttcgagctgtgaacacacactgtgtgaattctc 1211
Db 298 gtagctggtggaatgacacacatgttcgagctgtgaacacacactgtgtgaattctc 357
Oy 1212 tctagaagatctctctggaacacatattcagagaagatttctgtgtgtatctgcagt 1271
Db 358 tctagaagatctctctggaacacatattcagagaagatttctgtgtgtatctgcagt 417
Oy 1272 taactgtcttggaaacatttctgtcatattgcatgacgacctatatacagctcgtgaacaa 1331
Db 418 taactgtcttggaaacatttctgtcatattgcatgacgacctatatacagctcgtgaacaa 477
Oy 1332 gctgtatgcatgtcaatcattctctctgtctgtcgcagctgtgaatggaataattt 1391
Db 478 gctgtatgcatgtcaatcattctctctgtctgtcgcagctgtgaatggaataattt 537
Oy 1392 attcgtatcggaagcttggacctaagtttcgttgaagaatacaataaagacatgcgcagc- 1450
Db 538 attcgtatcggaagcttggacctaagtttcgttgaagaatacaataaagacatgcgcagc 597
Oy 1451 tctgtatggaagaatcattcattgtcagctgtga--ggaatccttggacattcgtccacaga 1508
Db 598 tctgtatggaagaatcattcattgtcagctgtga--ggaatccttggacattcgtccacaga 657
Oy 1509 agtaccagtttactgcttaacattctgcacattctggaacaaataacatcctgcatgtctacc 1568
Db 658 agtaccagtttactgctgtaacantctgcacattctggaacaaataacatcctgcatgtccacc 717
Oy 1569 tttagatgtgtgaga 1584
Db 718 nttagaaggaggaggaga 733

```

RESULT 7
 AAA44932
 ID AAA44932 standard; cDNA; 636 BP.
 XX
 AC AAA44932;
 XX
 DT 21-AUG-2000 (first entry)
 XX
 DE Human secreted expressed sequence tag SEQ ID NO:1507.
 XX
 KW Human; mouse; chicken; rat; secreted expressed sequence tag; SEST; expressed sequence tag; EST; probe; chemotactic; proliferative; immunomodulatory; haematopoietic; chemokine; analgesic; haemostatic; thrombolytic; antiinflammatory; cytostatic; antidiabetic; antifungal; antiviral; antidiabetic; antisthmatic; vulnary; antiparkinsonian; antitumor; osteoprotective; neuroprotective; neurotropic; antiparasitic; cerebroprotective; anticonvulsant; antidepressant; gene therapy; vaccine; autoimmune disorder; multiple sclerosis; allergic condition; insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer; lymphoid cell deficiency; burn; osteoporosis; osteoarthritis; central nervous system disorder; Alzheimer's disease; stroke; Parkinson's disease; Huntington's disease; coagulation disorder; haemophilia; thrombosis; inflammatory disorder; Crohn's disease; tumour; infection; depression; psoriasis; ss.
 KW
 OS Homo sapiens.
 XX
 PN WO200021991-A1.
 XX
 PD 20-APR-2000.
 XX
 PF 15-OCT-1999; 99WO-US24206.
 XX

```

PR 15-OCT-1998; 98US-0104436.
XX
PA (GENY ) GENETICS INST INC.
XX
PI Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
XX Merberg D, Treacy M, Bowman MR;
XX
DR WPI: 2000-317938/27.
XX
XX Isolated polynucleotides, and encoded proteins, comprising secreted
PT expressed sequence tags (SESTs), useful for treating various disorders
PT such as autoimmune, infectious, and central nervous system disorders -
XX
PS Claim 1; Page 593; 803pp; English.
XX
XX AAA43426 to AAA5925 represent specifically claimed secreted expressed
CC sequence tags (SESTs), isolated from human, mouse, chicken and rat
CC tissue sources. The SESTs can have a range of activities depending on
CC the tissues they were isolated from. The activities include:
CC chemotactic; proliferative; immunomodulatory; haematopoietic;
CC chemokine; analgesic; haemostatic; thrombolytic; antidiabetic;
CC cyostatic; antibacterial; antifungal; antiviral; antidiabetic;
CC antisthmatic; vulnary; antitumor; osteoprotective;
CC neurotropic; antiparkinsonian; antiparasitic; cerebroprotective;
CC anticonvulsant; and antidepressant. The SESTs can be used for gene
CC therapy and in vaccines. The SESTs are useful as probes for the
CC identification and isolation of full-length cDNAs and genomic DNA
CC molecules which correspond to the SESTs. Proteins encoded by the SESTs
CC are useful in assays for determining biological activity and raising
CC antibodies. They may be useful for treatment of autoimmune disorders
CC (multiple sclerosis, insulin dependent diabetes), allergic conditions
CC (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers,
CC osteoporosis, osteoarthritis, central nervous system disorders
CC (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation
CC disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's
CC disease), tumours, bacterial, fungal or viral infections, depression and
CC psoriasis. AAA45926 to AAA5931 represent linker variants which are given
CC in the exemplification of the present invention.
XX
SQ Sequence 636 BP; 217 A; 115 C; 112 G; 186 T; 6 other:

```

Query Match 16.2%; Score 580.4; DB 21; Length 636;
 Best Local Similarity 98.7%; Pred. No. 9.3e-126;
 Matches 603; Conservative 0; Mismatches 6; Indels 2; Gaps 2;

```

Oy 2016 gattttattctgcccattacaagctgttggaacccaattctctactctgaccacaag 2075
Db 22 gattttattctgcccattacaagctgttggaacccaattctctactctgaccacaag 81
Oy 2076 accatttaagaagaatgattcattcgtttgtgtatatactacagaacagaataatct-attg 2134
Db 82 accatttaagaagaatgattcattcgtttgtgtatatactacagaacagaataatctnnhgg 141
Oy 2135 aagaagaagtgtaagaacacatagctcattcattcattcattcattcattcattcattcattc 2194
Db 142 aagaagaagtgtaagaacacatagctcattcattcattcattcattcattcattcattcattc 201
Oy 2195 aggaatgcccactgagtaagaagccggaacctttcacacacacccctgtgaagtgac 2254
Db 202 aggaatgcccactgagtaagaagccggaacctttcacacacacccctgtgaagtgac 261
Oy 2255 tgatttccaatcaacagagactcaattcctatcaatgactgactgtaattcattctt 2314
Db 262 tgatttccaatcaacagagactcaattcctatcaatgactgactgtaattcattctt 321
Oy 2315 cgcagaagaatctgtgggggtgcttcacagagagattcgttgaagaagaataacacac 2374
Db 322 cgcagaagaatctgtgggggtgcttcacagagagattcgttgaagaagaataacacac 380
Oy 2375 aaatataattataatagctaaagataatatttcaagaagacatgaggaataataa 2434
Db 381 aaatataattataatagctaaagataatatttcaagaagacatgaggaataataa 440

```

QY 2435 aatgactaatgctcttacaagaaggaaatattatatacaataatgtatatattagtaga 2494
|||||
Db 441 aatgactaatgctcttacaagaaggaaatattatatacaataatgtatatattagtaga 500
QY 2495 catttgcataagaataaagaagaatctacttcagtagaacatcattcattttcttaaca 2554
|||||
Db 501 catttgcataagaataaagaagaatctacttcagtagaacatcattcattttcttaaca 560
QY 2555 tgcattattagtaaccacactatgtgcataagcattgcataatagtcctcggaagttaga 2614
|||||
Db 561 tgcattattagtaaccacactatgtgcataagcattgcataatagtcctcggaagttaga 620
QY 2615 cagtcagcaaac 2625
|||||
Db 621 cagtcagcaaac 631

RESULT 8
AAI9584
ID AAI9584 standard; cDNA: 530 BP.
XX
AC AAI9584;
XX
XX 04-JAN-2002 (first entry)
XX
DE Human expressed polynucleotide SEQ ID NO 47.
XX
KW Human; nootropic; neuroprotective; cytosolic; dermatological; virucide;
KW immunosuppressive; antihistaminic; anti-HIV; antibacterial; vulnery;
KW antiparkinsonian; antispasmodic; antianemic; antiarthritic; cancer;
KW antihypertensive; hepatoprotective; cerebroprotective; antiinflammatory;
KW antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;
KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine;
KW ss.
XX
XX Homo sapiens.
OS
XX
PN WO20015387-A1.
XX
PD 02-AUG-2001.
XX
PE 17-JAN-2001; 2001WO-US01310.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 02-MAR-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218280.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220963.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.

PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227909.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0234984.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235836.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.

PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PI Rosen CA, Barash SC, Ruben SM;
XX
XX WPI: 2001-465573/50.
DR P-PSDB: AAM99972.
XX
XX
PT Isolated digestive system associated polypeptide for treating,
PT preventing and/or prognosing disorders related to the digestive system
PT including digestive system cancers and also for testing and detection
PT e.g. diagnosis -
XX
XX
PS Claim 1; SEQ ID NO 47; 509bp + Sequence Listing; English.
XX
XX The invention relates to novel genes (AAI99548-AAI99604) and proteins
CC (AAM99936-AA99984) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC (d) wound healing; (e) neurological diseases such as cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 530 BP; 158 A; 105 C; 95 G; 171 T; 1 other;

Query Match 14.7%; Score 528.6; DB 22; Length 530;

Best Local Similarity 99.8%; Pred. No. 1.1e-113;
Matches 528; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1744 gccacagattatcagtcggaattttctctgtgtatattgtgcgcattatcatcata 1803
DB 2 gccacagattatcagtcggaattttctctgtgtatattgtgcgcattatcatcata 61
QY 1804 gttttccatggaagcagttttatagtcgttcatcaaaagtgccataagcaacgaa 1863
DB 62 gttttccatggaagcagttttatagtcgttcatcaaaagtgccataagcaacgaa 121
QY 1864 atacgaatcaagtaaaagagatgatccttgcacaaagttttcttatagta 1923
DB 122 atacgaatcaagtaaaagagatgatccttgcacaaagttttcttatagta 181
QY 1924 actgatgcattatgctgataccattttgtagtgaaatttcttactgcgtcagta 1983
DB 182 actgatgcattatgctgataccattttgtagtgaaatttcttactgcgtcagta 241
QY 1984 gaataccaggtacataaccttgggtagtgattttatctgcacataacagtgct 2043
DB 242 gaataccaggtacataaccttgggtagtgattttatctgcacataacagtgct 301
QY 2044 ttgaaccaattctctactctgacacaaagacatttaagaatgatcatcgctt 2103
DB 302 ttgaaccaattctctactctgacacaaagacatttaagaatgatcatcgctt 361
QY 2104 tggataactacagacaagaanaatctatgacagcaaaagtcagaaacatatgctca 2163
DB 362 tggataactacagacaagaanaatctatgacagcaaaagtcagaaacatatgctca 421
QY 2164 tcattcctcgggtggaagtgtgcccctgcagaagatgcaccctggatgaataacgcg 2223
DB 422 tcattcctcgggtggaagtgtgcccctgcagaagatgcaccctggatgaataacgcg 481
QY 2224 gacctttcacataccctcgtgaaatgtcacatgatttctcaatcaaga 2272
DB 482 gacctttcacataccctcgtgaaatgtcacatgatttctcaatcaaga 530
RESULT 9
AAL35644
ID AAL35644 standard; cDNA; 530 BP.
XX
AC AAL35644;
XX
DT 08-JAN-2002 (first entry)
XX
DE Human musculoskeletal system related polynucleotide SEQ ID NO 986.
XX
KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
KW vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein;
KW musculoskeletal system; ss.
XX
OS Homo sapiens.
XX
PN W0200155367-A1.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-0501338.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.

PR 25-SEP-2000; 2000US-234997P.
PR 25-SEP-2000; 2000US-234998P.
PR 26-SEP-2000; 2000US-235484P.
PR 27-SEP-2000; 2000US-235834P.
PR 27-SEP-2000; 2000US-235836P.
PR 29-SEP-2000; 2000US-236327P.
PR 29-SEP-2000; 2000US-236367P.
PR 29-SEP-2000; 2000US-236368P.
PR 29-SEP-2000; 2000US-236369P.
PR 29-SEP-2000; 2000US-236370P.
PR 02-OCT-2000; 2000US-236802P.
PR 02-OCT-2000; 2000US-237037P.
PR 02-OCT-2000; 2000US-237038P.
PR 02-OCT-2000; 2000US-237039P.
PR 02-OCT-2000; 2000US-237040P.
PR 13-OCT-2000; 2000US-239935P.
PR 13-OCT-2000; 2000US-239937P.
PR 20-OCT-2000; 2000US-240960P.
PR 20-OCT-2000; 2000US-241221P.
PR 20-OCT-2000; 2000US-241785P.
PR 20-OCT-2000; 2000US-241786P.
PR 20-OCT-2000; 2000US-241787P.
PR 20-OCT-2000; 2000US-241808P.
PR 20-OCT-2000; 2000US-241809P.
PR 20-OCT-2000; 2000US-241826P.
PR 01-NOV-2000; 2000US-244617P.
PR 08-NOV-2000; 2000US-246474P.
PR 08-NOV-2000; 2000US-246475P.
PR 08-NOV-2000; 2000US-246476P.
PR 08-NOV-2000; 2000US-246477P.
PR 08-NOV-2000; 2000US-246478P.
PR 08-NOV-2000; 2000US-246523P.
PR 08-NOV-2000; 2000US-246524P.
PR 08-NOV-2000; 2000US-246525P.
PR 08-NOV-2000; 2000US-246526P.
PR 08-NOV-2000; 2000US-246527P.
PR 08-NOV-2000; 2000US-246528P.
PR 08-NOV-2000; 2000US-246529P.
PR 08-NOV-2000; 2000US-246532P.
PR 08-NOV-2000; 2000US-246532P.
PR 08-NOV-2000; 2000US-246609P.
PR 08-NOV-2000; 2000US-246610P.
PR 08-NOV-2000; 2000US-246611P.
PR 08-NOV-2000; 2000US-246613P.
PR 17-NOV-2000; 2000US-249207P.
PR 17-NOV-2000; 2000US-249208P.
PR 17-NOV-2000; 2000US-249209P.
PR 17-NOV-2000; 2000US-249210P.
PR 17-NOV-2000; 2000US-249211P.
PR 17-NOV-2000; 2000US-249212P.
PR 17-NOV-2000; 2000US-249213P.
PR 17-NOV-2000; 2000US-249214P.
PR 17-NOV-2000; 2000US-249215P.
PR 17-NOV-2000; 2000US-249216P.
PR 17-NOV-2000; 2000US-249217P.
PR 17-NOV-2000; 2000US-249218P.
PR 17-NOV-2000; 2000US-249244P.
PR 17-NOV-2000; 2000US-249245P.
PR 17-NOV-2000; 2000US-249264P.
PR 17-NOV-2000; 2000US-249265P.
PR 17-NOV-2000; 2000US-249297P.
PR 17-NOV-2000; 2000US-249299P.
PR 17-NOV-2000; 2000US-249300P.
PR 01-DEC-2000; 2000US-250160P.
PR 01-DEC-2000; 2000US-250391P.
PR 05-DEC-2000; 2000US-251030P.
PR 05-DEC-2000; 2000US-251988P.
PR 05-DEC-2000; 2000US-256719P.
PR 06-DEC-2000; 2000US-251479P.
PR 08-DEC-2000; 2000US-251856P.
PR 08-DEC-2000; 2000US-251868P.
PR 08-DEC-2000; 2000US-251869P.
PR 08-DEC-2000; 2000US-251989P.
PR 08-DEC-2000; 2000US-251990P.
PR 11-DEC-2000; 2000US-254097P.

PR 05-JAN-2001; 2001US-259678P.
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
XX MPI: 2001-476161/51.
DR P-PSDB; ABB10249.
XX
XX Isolated nucleic acid molecule encoding an inflammation-associated
PT polypeptide is used in preventing, treating or ameliorating a medical
PT condition
PS Claim 1; SEQ ID NO: 137; 859pp + Sequence Listing; English.
XX
XX The present invention provides human cDNAs, proteins and related genomic
CC DNAs. These can be used in the treatment of neural, immune system,
CC muscular, reproductive, gastrointestinal, pulmonary, cardiovascular,
CC renal and proliferative disorders and inflammation. The present sequence
CC is a cDNA of the invention.
XX
XX Sequence 530 BP; 158 A; 105 C; 95 G; 171 T; 1 other:
SQ

Query Match 14.7%; Score 528.6; DB 22; Length 530;

Best Local Similarity 99.8%; Pred. No. 1.1e-113;

Matches 528; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1744 gccacattatcagtcggaatttctgtgattaatgtggcgcattatcacata 1803
DB 2 gccacattatcagtcggaatttctgtgattaatgtggcgcattatcacata 61
QY 1804 gtttttcctatggaagcagtggtttatagtgatcatcaaaagtcgcaaacactgaa 1863
DB 62 gtttttcctatggaagcagtggtttatagtgatcatcaaaagtcgcaaacactgaa 121
QY 1864 atacgaatcaagtttaaaaaagagatgacccctgccaagaagttttctttagratit 1923
DB 122 atacgaatcaagtttaaaaaagagatgacccctgccaagaagttttctttagratit 181
QY 1924 actgatgcattatgcctggaataccattttgtatgtaaatcttccactgcttcagta 1983
DB 182 actgatgcattatgcctggaataccattttgtatgtaaatcttccactgcttcagta 241
QY 1984 gaaataccaagttacccaataaccttgggtagtgatatttattctgcaccataacagtgc 2043
DB 242 gaaataccaagttacccaataaccttgggtagtgatatttattctgcaccataacagtgc 301
QY 2044 ttgaacccaattctctatctactgcacacaagaacatttaagaatgattcatcggtt 2103
DB 302 ttgaacccaattctctatctactgcacacaagaacatttaagaatgattcatcggtt 361
QY 2104 tggataactacagacaagaataatctatgagacagacaaggttcagaatacatatgctca 2163
DB 362 tggataactacagacaagaataatctatgagacagagcagtcaggaatacatatgctca 421
QY 2164 tcatlcatctgggtggaatggtggccactgvcaggaagatgccaactgagtaatgaagccg 2223
DB 422 tcatlcatctgggtggaatggtggccactgvcaggaagatgccaactgagtaatgaagccg 481
QY 2224 gacctttcacataccctctggaatgtaactgatttccataccaaga 2272
DB 482 gacctttcacataccctctggaatgtaactgatttccataccaaga 530
RESULT 11
ID AAS28950 standard; cDNA: 530 BP.
XX AAS28950;
AC AAS28950;
XX
DT 21-NOV-2001 (first entry)
XX

DE cDNA encoding for human uterine motility-association polypeptide #15.
XX
KW Human; uterine motility-association disorder; uterus; pregnancy;
KM labour; menstrual cycle; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN W0200155201-A1.
PD
XX 02-AUG-2001.
PF 17-JAN-2001; 2001WO-US01317.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 06-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.

PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI: 2001-488777/53.
XX
DR P-PSDB: AAU18108.
XX
PT Isolated polypeptide and nucleic acid molecules for treating,
PT preventing and/or prognosing disorders related to uterine motility
PT e.g. disorders associated with pregnancy and the menstrual cycle -
XX
PS Claim 4; SEQ ID No 25; 524pp; English.
XX
CC The present invention relates to the isolation of novel human
CC uterine motility-association polypeptides (AAU18094-AAU18152),
CC and cDNA and genomic sequences encoding for these polypeptides.
CC The sequences of the invention are useful in the diagnosis,
CC treatment, prevention and/or prognosis of diseases associated
CC with uterine motility such as pregnancy and labour, and menstrual
CC disorders. The polynucleotide sequences of the invention are also
CC useful in gene therapy. AAS28936-AAS28994 represent cDNA sequences
CC encoding for novel human uterine motility-association polypeptides.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX
SQ Sequence 530 BP; 158 A; 105 C; 95 G; 171 T; 1 other;

Query Match 14.7%; Score 528.6; DB 22; Length 530;
Best Local Similarity 99.8%; Pred. NO. 1.1e-113;
Matches 528; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1744 gccacagattatcagtggaattttctgtgataatttgccgcatattatcata 1803
DB 2 gccagattatcagtggaattttctgtgataatttgccgcatattatcata 61
QY 1804 gtttttctcatggaagcagttttatagtgatcacaagaatgcacaaagcagc 1863
DB 62 gtttttctcatggaagcagttttatagtgatcacaagaatgcacaaagcagc 121
QY 1864 atacggaatcaagttaaaaagagatgatccttgccaaagctttttcttataag 1923
DB 122 atacggaatcaagttaaaaagagatgatccttgccaaagctttttcttataag 181
QY 1924 actgatgcatatgctggaataccattttgtatggaattttcttactgctcag 1983
DB 182 actgatgcatatgctggaataccattttgtatggaattttcttactgctcag 241
QY 1984 gaaatccagtgctacataactcttggtatgataatttttttgcacataacag 2043
DB 242 gaaatccagtgctacataactcttggtatgataatttttttgcacataacag 301
QY 2044 ttgaacccaattctctatactcgaccacaagccatttaagaatgattcagttt 2103
DB 302 ttgaacccaattctctatactcgaccacaagccatttaagaatgattcagttt 361
QY 2104 tggataactacagacaaagaataatctatgacagcaaaagtgacagaacata 2163
DB 362 tggataactacagacaaagaataatctatgacagcaaaagtgacagaacata 421
QY 2164 tcatcatctgggtggaatagtgccacgcaagagatgacacacagtgattgaag 2223
DB 422 tcatcatctgggtggaatagtgccacgcaagagatgacacacagtgattgaag 481
QY 2224 gacctttcacataccctgtgaatgatactatttctcaatcaaga 2272
DB 482 gacctttcacataccctgtgaatgatactatttctcaatcaaga 530

RESULT 12
AAS29573
ID AAS29573 standard; cDNA; 530 BP.
XX
AC AAS29573;
XX
DT 21-NOV-2001 (first entry)
XX
DE Human endocrine polypeptide encoding cDNA SEQ ID No 73.
XX
KW Endocrine protein; human; mouse; rabbit; goat; horse; food additive;
KW cat; dog; chicken; sheep; immunosuppressive; antiarrhythmic; vasotropic;
KW antihemetic; antiproliferative; cytostatic; cardiant; neuroprotective;
KW cerebroprotective; nootropic; antibacterial; virucide; fungicide; cancer;
KW ophthalmological; vulnereary; gene therapy; autoimmune disease; neoplasm;
KW hyperproliferative disorder; breast; liver; cardiovascular disorder; ss;
KW cerebrovascular disorder; nervous system disorder; bacterial infection;
KW fungal infection; viral infection; ocular disorder; endocrine disorder;
KW gastrointestinal disorder; renal disorder; respiratory disorder;
KW wound healing; skin aging; organ transplantation; food preservative;
KW tissue regeneration; anti-fertility.
XX
OS Homo sapiens.
XX
PN MO200155364-A2.
XX
PD 02-AUG-2001.
XX
PE 17-JAN-2001; 2001MO-US01308.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189674.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.

OY 1924 actgatgcattatgctggtaccattttgtagtgaattcttccactgcagta 1983
|||||
Db 182 actgagcattatgctggtaccattttgtagtgaattcttccactgcagta 241
1984 gaaataccaggtaccataaccttggtagtattttattctgcacataacagtct 2043
|||||
Db 242 gaaataccaggtaccataaccttggtagtattttattctgcacataacagtct 301
OY 2044 ttgaacccaattctctatctctgaccacagaacccaatttaaaaatgatctggtt 2103
|||||
Db 302 ttgaacccaattctctatctctgaccacagaacccaatttaaaaatgatctggtt 361
OY 2104 ttgtataactacagacaaaagaatactatgacagcaaaagtcagaacatctgtcca 2163
|||||
Db 362 ttgtataactacagacaaaagaatactatgacagcaaaagtcagaacatctgtcca 421
OY 2164 tcatcatctggtgtagaattgtgcactgcaagagatgcaccctgagtaatgaagcgg 2223
|||||
Db 422 tcatcatctggtgtagaattgtgcactgcaagagatgcaccctgagtaatgaagcgg 481
OY 2224 gacctttcacataccctgtgaaatgtaactgactatttccaatcaaga 2272
|||||
Db 482 gacctttcacataccctgtgaaatgtaactgactatttccaatcaaga 530

RESULT 13

AA30187
ID AA30187 standard; cDNA; 530 BP.
XX
AC AA30187;
XX
DT 21-NOV-2001 (first entry)
XX
DE DNA encoding renal and cardiovascular-associated protein, Seq ID 33.
XX
KW Human; antiinflammatory; neuroprotective; immunomodulator; vulnereary;
cardiovascular; cytosolic; nephrotropic; antianaemic; nephritis;
immunosuppressive; kidney disorder; renal failure; hypertension;
cardiovascular disorder; myocardial infarction; blood disorder; anaemia;
blood coagulation disorder; electrolyte imbalance disorder; cancer;
hypnatraemia; hyperkalaemia; neoplastic disorder; nephroma;
autoimmune disease; inflammatory disease; reproductive system disorder;
endocrine disorder; neural activity; neurological disorder;
wound healing; respiratory disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200155328-A2.
PD
XX
XX 02-AUG-2001.
PF 17-JAN-2001; 2001WO-US01359.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.

PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 23-AUG-2000; 2000US-0227182.
PR 30-AUG-2000; 2000US-0227824.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234597.
PR 25-SEP-2000; 2000US-0234598.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239335.
PR 13-OCT-2000; 2000US-0239337.
PR 20-OCT-2000; 2000US-0240360.
PR 20-OCT-2000; 2000US-0241121.
PR 20-OCT-2000; 2000US-0241185.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.

PR	08-NOV-2000	2000US-0246478.
PR	08-NOV-2000	2000US-0246523.
PR	08-NOV-2000	2000US-0246524.
PR	08-NOV-2000	2000US-0246525.
PR	08-NOV-2000	2000US-0246526.
PR	08-NOV-2000	2000US-0246527.
PR	08-NOV-2000	2000US-0246528.
PR	08-NOV-2000	2000US-0246532.
PR	08-NOV-2000	2000US-0246609.
PR	08-NOV-2000	2000US-0246610.
PR	08-NOV-2000	2000US-0246611.
PR	17-NOV-2000	2000US-0246613.
PR	17-NOV-2000	2000US-0249207.
PR	17-NOV-2000	2000US-0249208.
PR	17-NOV-2000	2000US-0249209.
PR	17-NOV-2000	2000US-0249210.
PR	17-NOV-2000	2000US-0249211.
PR	17-NOV-2000	2000US-0249212.
PR	17-NOV-2000	2000US-0249213.
PR	17-NOV-2000	2000US-0249214.
PR	17-NOV-2000	2000US-0249215.
PR	17-NOV-2000	2000US-0249216.
PR	17-NOV-2000	2000US-0249217.
PR	17-NOV-2000	2000US-0249218.
PR	17-NOV-2000	2000US-0249244.
PR	17-NOV-2000	2000US-0249245.
PR	17-NOV-2000	2000US-0249264.
PR	17-NOV-2000	2000US-0249265.
PR	17-NOV-2000	2000US-0249297.
PR	17-NOV-2000	2000US-0249299.
PR	17-NOV-2000	2000US-0249300.
PR	01-DEC-2000	2000US-0250160.
PR	01-DEC-2000	2000US-0250391.
PR	05-DEC-2000	2000US-0251030.
PR	05-DEC-2000	2000US-0251988.
PR	05-DEC-2000	2000US-0256719.
PR	06-DEC-2000	2000US-0251479.
PR	08-DEC-2000	2000US-0251856.
PR	08-DEC-2000	2000US-0251868.
PR	08-DEC-2000	2000US-0251869.
PR	08-DEC-2000	2000US-0251989.
PR	08-DEC-2000	2000US-0251990.
PR	11-DEC-2000	2000US-0254097.
PR	05-JAN-2001	2001US-0259678.
XX		
PA	(HUMA-) HUMAN GENOME SCI INC.	
XA		
PI	Rosen CA, Barash SC, Ruben SM;	
PI	WPI; 2001-488787/53.	
XX	P-PSDB; AAU18666.	
DR		
XX		
PT	New polynucleotides and polypeptides, useful for diagnosing, treating,	
PT	preventing or prognosing e.g. kidney, cardiovascular, blood,	
PT	electrolyte imbalance or neoplastic disorders, autoimmune diseases,	
PT	cancers -	
PS	Claim 1; SEQ ID No 33; 506pp; English.	
XX		
XX		
CC	The invention relates to novel nucleic acids and polypeptides useful for	
CC	diagnosing, treating, preventing and/or prognosing disorders related to	
CC	these polypeptides. The polynucleotides are especially useful in the	
CC	diagnosis, prognosis, prevention and/or treatment of diseases which	
CC	include kidney disorders (e.g. renal failure or nephritis),	
CC	cardiovascular disorders (e.g. hypertension or myocardial infarction),	
CC	blood disorders (e.g. anaemia or blood coagulation disorders),	
CC	electrolyte imbalance disorders (e.g. hyponatraemia or hyperkalaemia),	
CC	neoplastic disorders (e.g. nephroma or renal cell cancer), autoimmune	
CC	diseases, cancers, inflammatory diseases, reproductive system	
CC	disorders, endocrine disorders, neural activity and neurological	
CC	disorders, wound healing and respiratory disorders. AAS30165-AAS30251	
CC	represent the novel human renal and cardiovascular-associated nucleic	
CC	acid sequences of the invention. Note: The sequence data for this patent	
CC		

CC did not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at:
CC [ftp.wipo.int/pub/published_pct_sequences](http://wipo.int/pub/published_pct_sequences).
CX
XX
SQ Sequence 530 BP; 158 A; 105 C; 95 G; 171 T; 1 other;

Query Match	14.7%;	Score 528.6;	DB 22;	Length 530;
Best Local Similarity	99.8%;	Pred. No. 1.1e-113;		
Matches 528; Conservative	1;	Mismatches 0;	Indels 0;	Gaps 0;

QY 1744 gccacagattatrcagtgccattttctctgtatttaattggccgcatrratcatcata 1803
|||||
pb 2 acccagattatrcaatgcacattttcttcgtatattatgtgccacattatcatcata 61
|||||

```

QY 1804 gttttttcctaagcagtgtttataagttgttcatacaagtgcataacaagcaactcgaa 1863
|||||
Db 62 gttttttcctaagcagtgtttataagttgttcatacaagtgccataacaagcaactcgaa 121

```

QY 1864 atacggaatcaagttataaaaagagagatgacctctgccaacgctttcttcttaagattt 1923
 |||||
 Db 122 atacggaatcaagttataaaaagagatgacctctgccaacgctttcttcttaagattt 181

Db 182 acgatgcatatgctgatacccatcttctgtagtgaattcttctaacgtcctcagta 241

Db 242 gaataccaggtaccataacctggtgtagtgaatttattctgcaccattaacagtgc 301

Db 302 ttgaaccaatctctatactctgacacaaagaccatttaagaatgatcatcggtt 361

Db 362 tgytatactacagacaagaatactatgacagcawggtcagaatacatatgtccca 421

Db 422 tcattcatctgtgtaaatgtgccaactgcagaagatgcaccctgagttaatgaagccg 481

Qy 2224 gacctttcaataccctcgtgaatcgtcactgatttcataatcaaga 2272

DB 482 gacccttccacataccctcgtgaatcgtcacatgcttctccaaatcagca 330

ID	AAS34845 standard; cDNA; 530 BP.
XX	
AC	AAS34845;

DT	04-DEC-2001	(first entry)
XX		
DE	cdna encoding novel human neoplastic disease associated polypeptide #77	

KM human neoplastic disease associated polypeptide, cancer, gene therapy
KM hyperproliferative disorder; neural disorder; immune system disorder;
KM muscular disorder; reproductive disorder; gastrointestinal disorder;
KM pulmonary disorder; cardiovascular disorder; renal disorder;
KM

XX Homo sapiens.
OS
XX

XX	
PD	02-AUG-2001.
XX	
DE	17-JUN-2001, 2001EO-11501358

PR 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235634.
PR 27-SEP-2000; 2000US-0235636.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.

PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239835.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0244674.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249266.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249300.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.
PA
XX
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
XX
DR WPI; 2001-465558/50.
DR P-PSDB; AUG21646.
XX
XX
PT Novel polypeptides and polynucleotides useful as diagnostic reagents to
PT diagnose diseases or disorders associated with aberrant expression or
PT activity of polypeptides, and for treating cancers, rheumatoid

PT arthritis -
XX
PS Claim 4; SEQ ID NO 89; 687pp; English.
XX
CC The present invention relates to the isolation of novel human neoplastic
CC disease associated polypeptides (AAU21568-AAU21561), and cDNA and DNA
CC sequences encoding for these polypeptides. The sequences of the
CC invention are useful in the diagnosis, treatment, prevention and/or
CC prognosis of disorders involving neoplastic disease such as
CC hyperproliferative disorders (e.g. leukemia, bone cancer, bladder
CC cancer, brain stem glioma, adult liver cancer, childhood cerebellar
CC astrocytoma, or Hodgkin's lymphoma). The sequences of the invention may
CC also be useful for treating other disorders such as neural disorders,
CC immune system disorders, muscular disorders, reproductive disorders,
CC gastrointestinal disorders, pulmonary disorders, cardiovascular disorders
CC and renal disorders. The polynucleotide sequences of the invention are
CC also useful in gene therapy. AAS34767-AAS35050 represent cDNA sequences
CC encoding for the novel human neoplastic disease associated polypeptides
CC of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pubd/published_pct_sequences.
XX
SQ Sequence 530 BP; 158 A; 105 C; 95 G; 171 T; 1 other;

Query Match 14.7%; Score 528.6; DB 22; Length 530;
Best Local Similarity 99.8%; Pred. No. 1.1e-113;
Matches 528; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1744 gccacagattatccagtcggaattctctgtgattattgtggcgcattatcatcata 1803
DB 2 gccacagattatccagtcggaattctctgtgattattgtggcgcattatcatcata 61

QY 1804 gtttttccctatggaagcagctgtttatagtgatcatcaaaagtgcatacagaacgtgaa 1863
DB 62 gtttttccctatggaagcagctgtttatagtgatcatcaaaagtgcatacagaacgtgaa 121

QY 1864 atcgcgaatcaagtttaaaagaagatgatccttcgccaaggttttttctttagtatt 1923
DB 122 atcgcgaatcaagtttaaaagaagatgatccttcgccaaggttttttctttagtatt 181

QY 1924 actgaatcatatgctcgtgataccattttgtatgtgaattcttctacgtctcagtt 1983
DB 182 actgaatcatatgctcgtgataccattttgtatgtgaattcttctacgtctcagtt 241

QY 1984 gaaatccaggtacataaactcttggtgtagtattttatcttgcattaaacgtgct 2043
DB 242 gaaatccaggtacataaactcttggtgtagtattttatcttgcattaaacgtgct 301

QY 2044 ttgaaccgaattctctatctctgacacagaacatttaaaagaatgattcgtgtt 2103
DB 302 ttgaaccgaattctctatctctgacacagaacatttaaaagaatgattcgtgtt 361

QY 2104 ttgtataactacagacaagaataatctatgacagcaaaagtgcagaanaacatagtccta 2163
DB 362 ttgtataactacagacaagaataatctatgacagcaaaagtgcagaanaacatagtccta 421

QY 2164 tcatcatctgggtggaatggtggtgcacgtgcagagagatgcacactgagttatgaagcgg 2223
DB 422 tcatcatctgggtggaatggtggtgcacgtgcagagagatgcacactgagttatgaagcgg 481

QY 2224 gacctttacatacccccgtgaaatgctcatttctcaatcaacga 2272
DB 482 gacctttacatacccccgtgaaatgctcatttctcaatcaacga 530

RESULT 15
ID AAL35575/C
XX AAL35575 standard; cDNA; 563 BP.
XX
AC AAL35575;
XX

DT 08-JAN-2002 (first entry)
XX
DE Human musculoskeletal system related polynucleotide SEQ ID NO 917.
XX
KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
XX antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
XX vulnerable; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein;
KW musculoskeletal system; ss.
XX
OS Homo sapiens.
XX
PN MO200155367-A1.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001MO-US01338.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225477.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227009.
PR 23-AUG-2000; 2000US-0227182.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.


```

Db 263 CTCCTTGCACTGCCAGCAGGAGATACTACTTTTAACTAGTTGTTCTTCCTAGTCT 204
QY 3415 ctacgttattagmatlitttgcttcataatgtaaaccttaagcaggagaagaaatg 3474
Db 203 CTACGTTATTAGAAATTTTGGCTTTCATTAATGTGAAACCTTTAAGCAGGAGAGAGAAATG 144
QY 3475 ttctcagatagttccaataacccaanaatgtttgcaacacacaaaatcactgaaac 3534
Db 143 TTTTCAGATAGTTTCAAAATACACCAAAAATGTTGAAACACAAAATTACTGAAATCAAAC 84
QY 3535 cataatgcccattatgaatatataatgtcgtatagnlttgctcgaana 3581
Db 83 CATPAATGCACCTTATTGAATATATAGTGTATATGATTTGTTCTGAAAA 37

```

Search completed: September 5, 2002, 08:24:32
 Job time: 16102 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 5, 2002, 02:59:25 : Search time 112.92 Seconds
(without alignments)
7796.232 Million cell updates/sec

Title: US-09-647-067-7

Perfect score: 3584
Sequence: 1 cgccttgcgaactgactgaaga.....tagntttgtctgaaaccc 3584

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 38353 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA: *
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq: *
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq: *
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq: *
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq: *
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq: *
6: /cgn2_6/ptodata/2/ina/Backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	59.8	1.7	3573	4	US-09-353-585-4
2	59.8	1.7	6471	4	US-09-353-585-1
3	58.8	1.6	2290	4	US-09-131-648-4
4	54.8	1.5	4417	4	US-07-741-453A-57
5	50.8	1.4	636	4	US-08-998-416-1137
6	50.6	1.4	2179	1	US-08-487-886-1
7	50.6	1.4	2179	2	US-08-531-070A-1
8	49.4	1.4	2179	3	US-08-483-855-1
9	49.4	1.4	837	4	US-08-998-416-288
10	47.6	1.3	2022	4	US-07-757-342D-1
11	47.6	1.3	2987	4	US-07-757-342D-1
12	47.6	1.3	3979	4	US-09-180-439-1
13	47.6	1.3	3979	4	US-09-180-439-2
14	47.6	1.3	4123	4	US-09-180-439-7
15	47.6	1.3	7218	4	US-08-233-463-14
16	46.8	1.3	3710	4	US-07-741-453A-62
17	46.8	1.3	8378	5	PCT-US91-09055-1
18	46.4	1.3	426	1	US-08-442-063A-35
19	46.4	1.3	567	1	US-08-442-063A-38
20	46.4	1.3	711	1	US-08-442-063A-41
21	46.4	1.3	849	1	US-08-442-063A-44
22	46.4	1.3	924	1	US-08-442-063A-47
23	46.4	1.3	1002	1	US-08-442-063A-26
24	46.4	1.3	1026	1	US-08-272-919-1
25	46.4	1.3	1026	1	US-08-619-916-1
26	46.4	1.3	1026	5	PCT-US95-08542-1
27	46.4	1.3	1593	6	5340934-3

28	44.2	1.2	4203	2	US-08-866-757-1	Sequence 1, Appl
29	44.2	1.2	4203	4	US-09-153-593-1	Sequence 1, Appl
30	44	1.2	2317	3	US-08-749-522-5	Sequence 5, Appl
31	44	1.2	3974	4	US-08-467-504-3	Sequence 3, Appl
32	43.8	1.2	2296	3	US-09-188-930-228	Sequence 28, App
33	43.6	1.2	615	4	US-08-998-416-186	Sequence 186, App
34	43.6	1.2	19124	2	US-08-487-826B-13	Sequence 13, Appl
35	43.4	1.2	2513	4	US-09-228-986-13	Sequence 13, Appl
36	43.4	1.2	51952	3	US-08-947-823-1	Sequence 1, Appl
37	43	1.2	4758	3	US-09-191-647-1	Sequence 1, Appl
38	43	1.2	4758	4	US-09-540-245A-1	Sequence 1, Appl
39	43	1.2	4758	4	US-09-540-153-1	Sequence 1, Appl
40	42.8	1.2	2168	3	US-08-749-522-6	Sequence 6, Appl
41	42	1.2	1677	2	US-08-684-101-1	Sequence 1, Appl
42	42	1.2	1677	4	US-09-205-814-1	Sequence 1, Appl
43	42	1.2	5176	4	US-09-182-024A-1	Sequence 1, Appl
44	41.6	1.2	3925	2	US-09-047-026A-3	Sequence 3, Appl
45	41.6	1.2	6854	1	US-08-468-036-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1
US-09-353-585-4
Sequence 4, Application US/09353585
Patent No. 6287865
GENERAL INFORMATION:
APPLICANT: Dixon, Mark S
Jones, Jonathan DG
TITLE OF INVENTION: Plant pathogen resistance genes and uses
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESS: Nixon & Vanderhye PC
STREET: 8th Floor, 1100 No. 6287865th Glebe Road
CITY: Arlington
STATE: Virginia
COUNTRY: United States of America
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/353,585
FILING DATE: 15-Jul-1999
CLASSIFICATION: C12N 15/29, 15/82, A01H 5/00, A01N 65/00, C12O 1/68
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/930,277
FILING DATE: 27-OCT-1997
APPLICATION NUMBER: PCT/GB96/00785
FILING DATE: 01-APR-1996
APPLICATION NUMBER: GB 9506558.5
FILING DATE: 31-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Ms Mary J Wilson
REGISTRATION NUMBER: 32,955
REFERENCE/DOCKET NUMBER: 620-69
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 3573 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO

Db 2562 GGAATTCAGCTTTCGCTCTATTCTCGAAGAAATAGTACTACCTAATCTTAAATGTC 2621
Qy 844 ttagatgagaggaagaaataataatcaacttaataatgaataactttgcaactctcag 903
Db 2622 CTAGGTTTGAGTGAAGATGCTTAAATGCTCTAATCTCTGCTTCATGAGGAAATCTGAAA 2681
Qy 904 aaactgagatgaattgattaggaagtaataagattgaataatctccacgccttatc 963
Db 2682 AACTGTGTAGTTCATCTTGTATATATACAGCTTTCGGCTCTAATCTCTGCTTCATG 2741
Qy 964 aaggaactgagagagctgacaaattgaactcttccataatc 1006
Db 2742 GGGATCTGACAACTGTCTATGTATGCTTTACATAAC 2784

RESULT 3
US-09-131-648-4
Sequence 4, Application US/09131648
Patent No. 6168920
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Yue, Henry
APPLICANT: Corley, Neil C.
APPLICANT: Guegler, Karl J.
APPLICANT: Patterson, Chandra
TITLE OF INVENTION: EXTRACELLULAR ADHESIVE PROTEINS
FILE REFERENCE: PF-0576 US
CURRENT APPLICATION NUMBER: US/09/131,648
CURRENT FILING DATE: 1998-08-10
NUMBER OF SEQ ID NOS: 5
SOFTWARE: PERL Program
SEQ ID NO 4
LENGTH: 2290
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: 2687731
US-09-131-648-4

Query Match 1.6%; Score 58.8; DB 4; Length 2290;
Best Local Similarity 47.5%; Pred. No. 3.2e-05;
Matches 174; Conservative 0; Mismatches 192; Indels 0; Gaps 0;

Qy 759 aagactacatggtcgtgaccttgaagcaaccataatccataatgaagaattgacttt 818
Db 338 aaacctactgctgactgattatccaacaattatcttcagtcaccaataatgatt 397
Qy 819 tattccgcgagtaatttaactgttttagtgatgaggaagaaacaaataataacttaa 878
Db 338 aaaaagatgcccagctccttctcgtgtaacctagaggaagaaacaaacttaactgacc 457
Qy 879 tgaataactcttgcacctctccagaacactgatatgaattgattaggaagtaataagat 938
Db 458 tgaataaagtctgctcgcagactgagcaacttaacaagaactatataatcaaaacttgct 517
Qy 939 tgaataactccacgcgttatataccaagaacctgaagagctgtccaacttgaatcttc 998
Db 518 ttctcaaatcttcaactgagccttatttgacctacataatcttctcgactcactcaaa 577
Qy 999 ctataatcaatccagaataatccaagcaaaccaattgatatcttgcacaaactaagtc 1058
Db 578 ttcaaatgattgcagatgatacaacagtaagtgtttagtgcctctccaactagagat 637
Qy 1059 tctcagcctagaagagattgaattccaataatccaacaaagagattgagacactcttat 1118
Db 638 tctgtagtctgggggaaatccaattatcagaatcaagaacagacatgaactttaagcctctat 697
Qy 1119 gaactc 1124
Db 698 caactc 703

RESULT 4
US-07-741-453A-57

Sequence 57, Application US/07741453A
Patent No. 6228597
GENERAL INFORMATION:
APPLICANT: PARMENTIER, MARC
APPLICANT: LIBERT, FREDERIC
APPLICANT: DUMONT, JACQUES
APPLICANT: VASSART, GILBERT
TITLE OF INVENTION: POLYPEPTIDES HAVING THYROTROPIN-RECEPTOR
TITLE OF INVENTION: ACTIVITY, NUCLEIC ACID SEQUENCES CODING FOR SUCH RECEPTORS
TITLE OF INVENTION: AND POLYPEPTIDES, AND APPLICATIONS OF THESE POLYPEPTIDES
NUMBER OF SEQUENCES: 62
CORRESPONDENCE ADDRESS:
ADDRESSEE: CUSHMAN DABRY & CUSHMAN
STREET: 1615 L STREET, N.W.
CITY: WASHINGTON, D.C.
COUNTRY: U.S.A.
ZIP: 20036
COMPUTER READABLE FORM:
MEDIUM TYPE: Disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/741,453A
FILING DATE: 19911015
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: KOKULIS, PAUL N.
REGISTRATION NUMBER: 16773
REFERENCE/DOCKET NUMBER: 91913/1107/US/ST
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 861-3000
TELEFAX: (202) 822-0944
TELEX: 6714627 CUSH
INFORMATION FOR SEQ ID NO: 57:
SEQUENCE CHARACTERISTICS:
LENGTH: 4417 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-07-741-453A-57

Query Match 1.5%; Score 54.8; DB 4; Length 4417;
Best Local Similarity 47.0%; Pred. No. 0.0004;
Matches 170; Conservative 0; Mismatches 192; Indels 0; Gaps 0;

Qy 1183 tgaacaacaacactgataagattcaatctcaggaatctcttgcaagcaattatcag 1242
Db 1254 TGTACTCCTAAGTCAAGATGAGTTCAACCCCTGTGAAGACATGATGGCTACAACTTCGCG 1313
Qy 1243 agagatttgcctggttgatctgacgtgactgcttggagaacattttgcaattggc 1302
Db 1314 AGGATTTGGTGTGGTGTGTGTTAGTCTGCTGGCTCTCTGGGGAAGTGTCTTCGATTC 1373
Qy 1303 atgcgacctatataagctctgagaacaagctgataatgcatgcaatctctctcgc 1362
Db 1374 GTCCTCCTTACAGACACAGTCAAAATTTGACGTGCCAGCGTTCTGATGCAACTTGGGC 1433
Qy 1363 tctgcgcagctcttaatggaatattatctgtagtcgagagcttgaactaaagttt 1422
Db 1434 TTTGCGAGATTTGCGATGGGGATGTATCTGCTCATCGCCTCCGTGACCTCTACACT 1493
Qy 1423 cgtggaagataaataaagatgcgacgctgtagatgagagagatcatcatgtgacttgta 1482
Db 1494 CATTCGTGATCTACAAACCATGCAATCCAGTGGCAAGAGCCCTGGGTGTACACAGCT 1553
Qy 1483 ggaactcttgccatctctccagaagaagatcatgatttacttgtaattctgacatg 1542
Db 1554 GGTTCCTTCACGTCTTGGCCAGATGATTAATACAGTGTATACAGTCAACGATCAACCGCG 1613

OY 1543 ga 1544
DB 1614 GA 1615

RESULT 5
US-08-998-416-1137/C
; Sequence 1137, Application US/08998416
; Patent No. 6239264

GENERAL INFORMATION:

APPLICANT: Philippsen, Peter
APPLICANT: Pohlmann, Rainer
APPLICANT: Steiner, Sabine
APPLICANT: Mohr, Christine
APPLICANT: Wendland, Jürgen
APPLICANT: Knechtle, Philipp
APPLICANT: Reibschung, Corinne
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPHII
TITLE OF INVENTION: AND USES THEREOF
NUMBER OF SEQUENCES: 1152
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6239264artis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: No. 6239264th Carolina
COUNTRY: USA
ZIP: 27709

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/998,416
FILING DATE: 24-DEC-1997
CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: CH 0016/97
FILING DATE: 31-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy

REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689

TELECOMMUNICATION INFORMATION:

INFORMATION FOR SEQ ID NO: 1137:
SEQUENCE CHARACTERISTICS:
LENGTH: 636 base pairs
TYPE: nucleic acid

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)

ORGANISM: PAG1692RP

US-08-998-416-1137

Query Match 1.4%; Score 50.8; DB 4; Length 636;
Best Local Similarity 48.08; Pred. No. 0.0018;
Matches 145; Conservative 0; Mismatches 157; Indels 0; Gaps 0;

OY 2274 acgaattccattcagactgactcgaatctatctctcgcagagaatactg99g 2333
DB 589 ATTAAATAATGATATAATATTAAATTAATTAATTAATTAATTAATTAATTA 530
OY 2334 gtgctcaatgaggattactactgtaataatgaataccacaataattataataat 2393
DB 529 AATGAATATATATTATTAATTAATTAATTAATTAATTAATTAATTAATTA 470
OY 2394 agctaagataataatttttaacaagacatgaggaataataataatgacttaata 2453

DB 469 ATTTAATATATATCTTATAAAGATTAATAATTAATTAATTAATTAATTAATAA 410

OY 2454 aagggaagtaataataataataataataataataataataataataataata 2513

DB 409 AATGATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 350

OY 2514 agagaatactactcagtaacatcatcattcttctaacaatgacattatgatacca 2573

DB 349 AATTAATATATATTTTATTAATAACATTAATTAATTAATTAATTAATTAATTA 290

OY 2574 ct 2575

DB 289 TT 288

RESULT 6
US-08-487-886-1

; Sequence 1, Application US/08487886
; Patent No. 5744448

GENERAL INFORMATION:

APPLICANT: Kelton, Christie Ann
APPLICANT: Schweickhardt, Rene Lynn
APPLICANT: Cheng, Shirley Vui Yen
APPLICANT: Nugent, No. 5744448een Patrice
TITLE OF INVENTION: Human Follicle Stimulating
TITLE OF INVENTION: Hormone Receptor
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Stephan P. Williams,
STREET: Ares-Serono, Inc.
CITY: Boston
STATE: MA

COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" diskette, 1.44 MB, high density
COMPUTER: IBM PS/2, model 55 SX
OPERATING SYSTEM: MS-DOS version 4.0
SOFTWARE: VAX/VMS Massll via Kermit to IBM MS-DOS

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/487,886
FILING DATE: 07-JUN-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/670,085
FILING DATE: 15-MAR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Williams, Stephan P.
REGISTRATION NUMBER: 28546
REFERENCE/DOCKET NUMBER: US/252

TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 723-1300
TELEFAX: (617) 723-8923

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2179

TYPE: Nucleic acid
STRANDEDNESS: Double
TOPOLOGY: linear

MOLECULE TYPE: cDNA to mRNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens

TISSUE TYPE: Testis
IMMEDIATE SOURCE:
LIBRARY: 19t11 cDNA library, Clontech #HL1010b

CLONE: pPHSR11-11, pPHSR15-6
FEATURE:
NAME/KEY: protein coding region
LOCATION: 75 to 2159

US-08-487-886-1


```

? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (617) 723-1300
? TELEFAX: (617) 723-8923
? INFORMATION FOR SRQ ID NO: 1:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 2179
? TYPE: Nucleic acid
? STRANDEDNESS: Double
? TOPOLOGY: Linear
? MOLECULE TYPE: cDNA to mRNA
? ORGANICLE SOURCE:
? ORGANISM: Homo sapiens
? TISSUE TYPE: Testis
? IMMEDIATE SOURCE:
? LIBRARY: IgT11 cDNA library, Clontech #HL1010b
? CLONE: pHRS11-11, pHRS15-6
? FEATURE:
? NAME/KEY: protein coding region
? LOCATION: 75 to 2159
? OS-08-482-855-1

```

Query Match	1.4%	Score 50.6	DB 3	Length 2179
Best Local Similarity	45.4%	Pred. No. 0.0033		
Matches 182, Conservative	0	Mismatches 219,	Indels 0	Gaps 0

QY	1149	ccagatctgaggatgacaccacacatttcgacgcctcttaaccacaaacctgagatctc	1208
Db	1076	CTATGACTTTTTCGAATGAAAGTGGTTACGTGACCTGCTCCCTTAAGCCAGATGACTTCAA	1135
QY	1209	atctctagagaactctcttggcaagacattatagaagagatctgtctggtgtglatctgc	1268
Db	1136	CCCATGTGAAGATATCATATGGGGATACAAACATCCTCAGAGTCCTGATATGGTTTATCAGAT	1195
QY	1269	agttacctgctttggaaacattttgtcaatttgacatgacacttatcagagctgtaa	1328
Db	1196	CTGTGGCCATCAGCTGGGAAACATCATAGTGGTACAGATCCTTAATACAGCCAAATATAACT	1255
QY	1329	caagctgtatgacatgtcaatacatctctctgcgtctgtgcagctgtcttaaggaaata	1388
Db	1256	CACAGTCCCAAGGTCCTTATGTGCAACCTGGCGCTTGTGTCTGTGATGGATGTAA	1315
QY	1389	tttatcgtatcggagagcttgaacctaaagttctgtgagatacaataagcatgagca	1448
Db	1316	CTGTGCTGCTATTTGCATTCATTTATATCCATACCAAGGCCAATATCACAACTATGCCAT	1375
QY	1449	gcctgtgagatgagatcactcattgtcaagcttgbagatcttggccattctgcacaga	1508
Db	1376	TGATGTGCACAAACTGGGGCAGGCTGTGATGCTGTGCTTTTCACTGTCCTTGGCCAGTGA	1435
QY	1509	agatcacgctttatctgttaacatttctgcacattgtgaaaaat	1549
Db	1436	GCCTGCAGTCTACACTGTGACACCTTTCACCTTTGGAAATAT	1476

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

CITY: Research Triangle Park
 STATE: No. 6239264th Carolina
 COUNTRY: USA
 ZIP: 27709
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.3
 CURRENT APPLICATION NUMBER: US/08/998,416
 FILING DATE: 24-DEC-1997
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: CH 0016/97
 FILING DATE: 31-DEC-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Meigs, J. Timothy
 REGISTRATION NUMBER: 36,241
 REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 919-541-8587
 TELEFAX: 919-541-8689
 INFORMATION FOR SEO ID NO.: 288:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 837 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 ORIGINAL SOURCE:
 ORGANISM: PAG1241RP
 US-08-998-416-288

Query Match	1.4%	Score 49.4	DB 4	Length 837
Best Local Similarly	50.6%	Pred. No. 0.0045		
Matches 119, Conservative	0	Mismatches 116,	Indels 0	Gaps 0

Oy	2341	atgagggacttactggtatgaaatagaataacccaanaaatlaattatataatgctag	2400
Db	522	ATAATTTATTAAATTAATATATAAAAAATTAATTAATAGAAAATTAAGTAAATTAATTTAA	463
Oy	2401	ataaatafttacagaagacatgagaaaataaaatgaactatgccttiacaaggaa	2460
Db	462	TAAATTAATCTTATTAAGAAGTTAAATRAATRTAAATCAACATTAATTTTATAAAAATGAAT	403
Oy	2461	gtaattatcaatbaatgatatalatatagtagacatttgcataagaataagaagaa	2520
Db	402	ATTATTAATTAATAAATTAATTTTCATTAATTTTAAATTAATTAATCGTTATTAATTAATAAA	343
Oy	2521	tctaactcagtaacatcatcatttttctaacaagcatitratagtaaccact	2575
Db	342	TATTTATTTTATTAACAACATTAATTAATTAATTAATTAATTAATGATATGCATTT	288

RESULT 10.
US-07-757-342D-9
Sequence 9, Application US/07757342D
Patent No. 6218509
GENERAL INFORMATION:
APPLICANT: IGARASHI, Masao
MINEGISHI, Takashi
NAKAMURA, Kazuo
TITLE OF INVENTION: PROTEIN, DNA AND USE THEREOF
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESSES:
ADDRESSEE: DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS &
CUSHMAN
STREET: 130 Water Street
CITY: Boston
STATE: Massachusetts
COUNTRY: US

```

1      ZIP: 02109
2
3      COMPUTER READABLE FORM:
4
5          MEDIUM TYPE: floppy disk
6
7      COMPUTER: IBM PC compatible
8
9      OPERATING SYSTEM: PC-DOS/MS-DOS
10     SOFTWARE: PatentIn Release #1.0, Version #1.25
11
12     CURRENT APPLICATION DATA:
13
14         APPLICATION NUMBER: US/07/757,342D
15
16         FILING DATE: 10-Sep-1991
17
18         CLASSIFICATION: <Unknown>
19
20     ATTORNEY/AGENT INFORMATION:
21
22         NAME: BUCKLEY, Linda M.
23
24         REGISTRATION NUMBER: 31003
25
26         REFERENCE/DOCKET NUMBER: 41226
27
28     TELECOMMUNICATION INFORMATION:
29
30         TELEPHONE: (617)523-3400
31
32         TELEFAX: (617)523-6440
33
34         TELEX: 200291 STRE UR
35
36     INFORMATION FOR SEQ ID NO: 9:
37
38         SEQUENCE CHARACTERISTICS:
39
40             LENGTH: 2022 base pairs
41
42             TYPE: nucleic acid
43
44             STRANDEDNESS: double
45
46             TOPOLOGY: linear
47
48         MOLECULE TYPE: cDNA to mRNA
49
50         FEATURE:
51
52             NAME/KEY: CDS
53
54             LOCATION: 1..2022
55
56         SEQUENCE DESCRIPTION: SEQ ID NO: 9:
57
58     JS-07-757-342D-9

```

```

Query Match      1.38; Score 47.6; DB 4; Length 2022;
Best Local Similarity 44.88; Pred. No. 0.018;
Matches 182; Conservative 0; Mismatches 224; Indels 0; Gaps 0;

QY 1167 accacatgctgcagctgttaaaccaacacacatgataatctcatctcagagaactctt 1226
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 936 ACCCAAGACACCCGATGTGCTCTCGAAGCACAGATCTTTAATCCCTGTGAAGACATTAT 995

QY 1227 ggcaagacatattcagagagatttgcctggctgtgtaatctgcagttacactgcttggaaa 1286
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 996 GGGGTATGACTTCCCTTAGGGTCCTGATTGGCTGATTAATATTTCAGGCATCATGGGAAA 1055

QY 1287 cattttctcaatttgcacagcgacctatatacagctctgagaaacaagctgtaagcatgctc 1346
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1056 CATGACTGTCTCTTTGTTGTCCTGACAAAGTCGTAAACACTTAAGAGTCCCTGTTTTCT 1115

QY 1347 aatcattctctgcgtctgcgcgaactgcttaaggaatatatttcgtatcggagag 1406
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1116 CATGTGCANATCTCTCTTTGGAGATTTTGCATGGGGCTCATTCGCTGCTATAGCCCTC 1175

QY 1407 cttagaacataagttctgtgagaatatacaataagaatcgacgctgtgagatgagagtaac 1466
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1176 AGTTGATTTCCCAACCAAGGCGCACTACTATTAAACCATGACATAGAGTGGCAAGACGGGAG 1235

QY 1467 tcaattgcagcttggagatcttttggccatctctgcacagaagaatatacgaatttactgtt 1526
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1236 TGGGTGCAGCACTGCTGGCTTTTACAGTATTTGCCAAGTAGAATCTTCTGTGTACACCCT 1295

QY 1527 aacatttctgacatgtaaaaaatacatatcgtcatctgtatcctttt 1572
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1296 CACCGTCATCACTCTAGAAAAGATGGCACACCATCAACCTATGTGTATT 1341

RESULT 11
US-07-757-342D-1
: Sequence 1, Application US/07757342D
: Patent No. 6218509
: GENERAL INFORMATION:
: APPLICANT: IGARASHI, Masao
: INVENTOR: IGARASHI, Masao
: MINEGISHI, Takashi
: NAKAMURA, Kazuo
:

```

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

	Query Match	1.3%	Score 47.6	DB 4	Length 2987
	Best Local Similarity	44.8%	Pred. No. 0.021		
	Matches 182	Conservative	0	Mismatches 224	Indels 0
					Gaps 0
Qy	1167	accacatgttcgcagctgttaaaccacaacacgcgatggaattcattcatctcagagaatcctt	1226		
Db	1011	AOCCAAGACACCCCGATGTGCTCCGGAACCAAGATCCTTTTAATCCCTGTGAAGACATTAAT	1070		
Qy	1227	ggcaagcaattatcagagagatttgtcttggtgtgtatctgcagttaccctgttgaaa	1286		
Db	1071	GGGCGTATACATCTCCCTAAGGGCTCGATTGGCTGATTAATATTTCAGGCATCATGGGAAA	1130		
Qy	1287	catcttggtcaatttcacatgcagacattatcagatgcgtcagacaagcgtatgcgatgc	1346		
Db	1131	CATGACTGTTCTTTTGTGTTCCTCTACAAAGTCGTTACAAACTTAAGTGCCCTCGTTTCT	1190		
Qy	1347	aatcaatttcctctctgtctgtgcgcactgtcttaatgggaatatattatctcgtatcgagag	1406		
Db	1191	CATGCGAATCTCTCCTTTGCGAGACTTTTGGCAATGGGGCTCTATCTGCTGCATGACCTTC	1250		
Qy	1407	ctttgaactaaagttctgttgagaaataaataagatgcgcagcgtgtgtagaggaagctac	1466		
Db	1251	AGTTGATTTCCCAACCAAGGGCGCACTACTATTAACCATGCCATAGACTGGCAGACGGAG	1310		
Qy	1467	tcaatgtcagcgttgaagatctcttggccatctctgtccacaagaagratcaatttactgt	1526		
Db	1311	TGGGCGACGACACTGCTGGCTTTTTCACGTATATGCCAAGTGAACCTTCTGTCTACACCT	1370		
Qy	1527	aacattctgcacatggaanaataacatcatcgtatgtctatccctt	1572		
Db	1371	CACCGTCATCACTGTAGAAGATGGCACACATCATCCATCTCTATT	1416		

```
RESULT 12
US-09-180-439-1
; Sequence 1, Application US/09180439
; Patent No. 6225532
; GENERAL INFORMATION:
; APPLICANT: Dixon, Mark S
; APPLICANT: Hatziathanis, Kostas
; APPLICANT: Jones, David A
; APPLICANT: Jones, Jonathan DG
; TITLE OF INVENTION: Plant pathogen resistance genes and uses thereof
; FILE REFERENCE: 620 - 53
; CURRENT APPLICATION NUMBER: US/09/180,439
; CURRENT FILING DATE: 1998-12-06
; EARLIER APPLICATION NUMBER: PCT/GB97/01249
; EARLIER FILING DATE: 1997-05-08
; EARLIER APPLICATION NUMBER: GB 9609681.3
; EARLIER FILING DATE: 1996-05-09
; EARLIER APPLICATION NUMBER: GB 9619924.5
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 3979
; TYPE: DNA
; ORGANISM: Lycopersicon esculentum
US-09-180-439-1

Query Match      1.3%; Score 47.6; DB 4; Length 3979;
Best Local Similarity 48.2%; Pred. No. 0.023;
Matches 134; Conservative 0; Mismatches 144; Indels 0; Gaps 0;

QY 444 tcttgatgtctcaagaattatcatatctctcaagaagctgagactgcaacaataagat 503
      ||||| || || |||| | | | | | | | | | | | | | | | | | | | | | |
DB 1778 tcttgatgtctcaagaattatcatatctctcaagaagctgagactgcaacaataagat 1837
      ||||| || || |||| | | | | | | | | | | | | | | | | | | | | | |
QY 504 tacatccatccatcatatgcttcgagagagctgaataagccttaactatgctatcag 563
      ||||| || || |||| | | | | | | | | | | | | | | | | | | | | | |
DB 1838 ttctgctctatctccgaagaataagttaccgtctctctcaagaactatattggg 1897
      ||||| || || |||| | | | | | | | | | | | | | | | | | | | | | |
QY 564 tcatacagaataacctctctgaagccgggtgttttgaagatcttccacagactagaatg 623
      ||||| || || |||| | | | | | | | | | | | | | | | | | | | | | |
DB 1898 taataactctcttaagctctatctctgcttcaattgggagactgaacaactgtttat 1957
      ||||| || || |||| | | | | | | | | | | | | | | | | | | | | | |
QY 624 gctgataattgaagaataacctcagctcagctgaattcccccacaactttatgactaa 683
      ||||| || || |||| | | | | | | | | | | | | | | | | | | | | | |
DB 1958 gtgtatcttcaacaatacagcttctgctctatctcctgaagaataagttactcag 2017
      ||||| || || |||| | | | | | | | | | | | | | | | | | | | | | |
QY 684 ttctctatctcttagtctcgtatgaataagctcctca 721
      ||||| || || |||| | | | | | | | | | | | | | | | | | | | | | |
DB 2018 ttctctactgaactattttgggtataaactctcta 2055
      ||||| || || |||| | | | | | | | | | | | | | | | | | | | | | |

RESULT 13
US-09-180-439-2
; Sequence 2, Application US/09180439
; Patent No. 6225532
; GENERAL INFORMATION:
; APPLICANT: Dixon, Mark S
; APPLICANT: Hatziathanis, Kostas
; APPLICANT: Jones, David A
; APPLICANT: Jones, Jonathan DG
; TITLE OF INVENTION: Plant pathogen resistance genes and uses thereof
; FILE REFERENCE: 620 - 53
; CURRENT APPLICATION NUMBER: US/09/180,439
; CURRENT FILING DATE: 1998-12-06
; EARLIER APPLICATION NUMBER: PCT/GB97/01249
; EARLIER FILING DATE: 1997-05-08
; EARLIER APPLICATION NUMBER: GB 9609681.3
; EARLIER FILING DATE: 1996-05-09
; EARLIER APPLICATION NUMBER: GB 9619924.5
; EARLIER FILING DATE: 1996-09-24
```

```
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 2
; LENGTH: 3979
; TYPE: DNA
; ORGANISM: Lycopersicon esculentum
US-09-180-439-2

Query Match      1.3%; Score 47.6; DB 4; Length 3979;
Best Local Similarity 48.2%; Pred. No. 0.023;
Matches 134; Conservative 0; Mismatches 144; Indels 0; Gaps 0;

QY 444 tcttgatgtctcaagaattatcatatctctcaagaagctgagactgcaacaataagat 503
      ||||| || || |||| | | | | | | | | | | | | | | | | | | | | | |
DB 1778 tcttgatgtctcaagaattatcatatctctcaagaagctgagactgcaacaataagat 1837
      ||||| || || |||| | | | | | | | | | | | | | | | | | | | | | |
QY 504 tacatccatccatcatatgcttcgagagagctgaataagccttaactatgctatcag 563
      ||||| || || |||| | | | | | | | | | | | | | | | | | | | | | |
DB 1838 ttctgctctatctccgaagaataagttaccgtctctctcaagaactatattggg 1897
      ||||| || || |||| | | | | | | | | | | | | | | | | | | | | | |
QY 564 tcatacagaataacctctctgaagccgggtgttttgaagatcttccacagactagaatg 623
      ||||| || || |||| | | | | | | | | | | | | | | | | | | | | | |
DB 1898 taataactctcttaagctctatctctgcttcaattgggagactgaacaactgtttat 1957
      ||||| || || |||| | | | | | | | | | | | | | | | | | | | | | |
QY 624 gctgataattgaagaataacctcagctcagctgaattcccccacaactttatgactaa 683
      ||||| || || |||| | | | | | | | | | | | | | | | | | | | | | |
DB 1958 gtgtatcttcaacaatacagcttctgctctatctcctgaagaataagttactcag 2017
      ||||| || || |||| | | | | | | | | | | | | | | | | | | | | | |
QY 684 ttctctatctcttagtctcgtatgaataagctcctca 721
      ||||| || || |||| | | | | | | | | | | | | | | | | | | | | | |
DB 2018 ttctctactgaactattttgggtataaactctcta 2055
      ||||| || || |||| | | | | | | | | | | | | | | | | | | | | | |

RESULT 14
US-09-180-439-7
; Sequence 7, Application US/09180439
; Patent No. 6225532
; GENERAL INFORMATION:
; APPLICANT: Dixon, Mark S
; APPLICANT: Hatziathanis, Kostas
; APPLICANT: Jones, David A
; APPLICANT: Jones, Jonathan DG
; TITLE OF INVENTION: Plant pathogen resistance genes and uses thereof
; FILE REFERENCE: 620 - 53
; CURRENT APPLICATION NUMBER: US/09/180,439
; CURRENT FILING DATE: 1998-12-06
; EARLIER APPLICATION NUMBER: PCT/GB97/01249
; EARLIER FILING DATE: 1997-05-08
; EARLIER APPLICATION NUMBER: GB 9609681.3
; EARLIER FILING DATE: 1996-05-09
; EARLIER APPLICATION NUMBER: GB 9619924.5
; EARLIER FILING DATE: 1996-09-24
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 7
; LENGTH: 4123
; TYPE: DNA
; ORGANISM: Lycopersicon esculentum
US-09-180-439-7

Query Match      1.3%; Score 47.6; DB 4; Length 4123;
Best Local Similarity 48.2%; Pred. No. 0.023;
Matches 134; Conservative 0; Mismatches 144; Indels 0; Gaps 0;

QY 444 tcttgatgtctcaagaattatcatatctctcaagaagctgagactgcaacaataagat 503
      ||||| || || |||| | | | | | | | | | | | | | | | | | | | | | |
DB 1922 tcttgatgtctcaagaattatcatatctctcaagaagctgagactgcaacaataagat 1981
      ||||| || || |||| | | | | | | | | | | | | | | | | | | | | | |
QY 504 tacatccatccatcatatgcttcgagagagctgaataagccttaactatgctatcag 563
      ||||| || || |||| | | | | | | | | | | | | | | | | | | | | | |
DB 1982 ttctgctctatctccgaagaataagttaccgtctctctcaagaactatattggg 2041
      ||||| || || |||| | | | | | | | | | | | | | | | | | | | | | |
```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2002, 05:27:20 ; Search time 61.35 Seconds
(without alignments)
1307.177 Million cell updates/sec

Title: US-09-647-067-8

Perfect score: 3834
Sequence: 1 MTSGVFFYILLFGKFFSHG.....FTYPCEMSLISOSTRLNYS 722

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

A-Geneseq_032802: *
1: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1980.DAT: *
2: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1981.DAT: *
3: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1982.DAT: *
4: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1983.DAT: *
5: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1984.DAT: *
6: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1985.DAT: *
7: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1986.DAT: *
8: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1987.DAT: *
9: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1988.DAT: *
10: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1989.DAT: *
11: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1990.DAT: *
12: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1991.DAT: *
13: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1992.DAT: *
14: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1993.DAT: *
15: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1994.DAT: *
16: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1995.DAT: *
17: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1996.DAT: *
18: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1997.DAT: *
19: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1998.DAT: *
20: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1999.DAT: *
21: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA2000.DAT: *
22: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA2001.DAT: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	3834	100.0	722	20	AA142171	Human IGR7 short f
2	3687.5	96.2	757	20	AA142170	Human IGR7 long fo
3	2091	54.5	396	22	AAE02498	Human CON222 G pro
4	2079	54.2	396	21	AA157286	Human GPCR protein
5	1082.5	28.2	355	22	AAU04370	Human G-protein co
6	948	24.7	188	22	AAAG9945	Human expressed po
7	912	23.8	176	22	AAAG9972	Human expressed po
8	912	23.8	176	22	ABB04062	Human expressed po
9	912	23.8	176	22	ABB10249	Human musculoskele
10	912	23.8	176	22	AAU18108	Human cDNA SEQ ID
11	912	23.8	176	22	AAU18344	Novel human uterin
						Human endocrine po

12	912	23.8	176	22	AAU18666	Renal and cardiova
13	912	23.8	176	22	AAU21646	Novel human neopla
14	717.5	18.7	172	22	ABB04105	Human musculoskele
15	650	17.0	140	21	AA841526	Human ORFX ORF1290
16	638	16.6	157	22	AA835407	Human HGR101 G-pr
17	599.5	15.6	359	22	ABB61216	Drosophila melanog
18	599.5	15.6	359	22	AAU38931	Drosophila G-prote
19	527	13.7	699	14	AA830517	N-terminal of LH r
20	511.5	13.3	692	14	AA830503	N-terminal of LH r
21	511.5	13.3	698	14	AA830505	N-terminal of LH r
22	511	13.3	699	14	AA830512	N-terminal of LH r
23	492.5	12.8	1300	22	ABB64083	Drosophila melanog
24	490.5	12.8	689	14	AA830509	N-terminal of LH r
25	489	12.8	699	14	AA830515	N-terminal of LH r
26	483.5	12.6	695	14	AA830524	N-terminal of LH r
27	482.5	12.6	695	14	AA830506	N-terminal of LH r
28	480	12.5	634	14	AA830520	N-terminal of LH r
29	480	12.5	693	14	AA830510	N-terminal of LH r
30	479.5	12.5	706	14	AA830504	N-terminal of LH r
31	479	12.5	685	12	AA81331	Human luteinizing
32	479	12.5	695	13	AA827358	FSHR. Homo sapien
33	478.5	12.5	692	11	AA808038	Rat testicular lut
34	478	12.5	695	14	AA842082	FSH receptor. Hom
35	478	12.5	695	18	AAW14782	FSH receptor. Hom
36	478	12.5	696	14	AA830523	N-terminal of LH r
37	477.5	12.5	334	22	ABB60463	Drosophila melanog
38	477.5	12.5	334	22	AAU8929	Drosophila G-prote
39	476.5	12.4	907	20	AAW3889	Human HG38 protein
40	476.5	12.4	907	21	AA90682	Human G-protein-co
41	476.5	12.4	907	21	AA90687	Human mutant G pro
42	476.5	12.4	951	20	AAV42168	Human IGR4 protein
43	474.5	12.4	700	14	AA830516	N-terminal of LH r
44	470.5	12.3	695	14	AA830525	N-terminal of LH r
45	470.5	12.3	700	11	AA808037	Rat ovarian lutein

ALIGNMENTS

RESULT 1
AA142171 standard; Protein: 722 AA.
ID
AC AAV42171;
XX
XX 20-DEC-1999 (first entry)
XX
XX Human IGR7 short form protein sequence.
DE
XX Human; IGR4; IGR5; IGR7: G-protein coupled receptor; gene therapy;
KW extracellular leucine rich repeat region; mapping; identification.
XX
XX Homo sapiens.
OS
XX
XX MO9948921-A1.
PN
XX 30-SEP-1999.
PD
XX
XX 25-MAR-1999; 99WC-US06573.
PF
XX
XX 26-MAR-1998; 98US-0079501.
PR
XX
XX (STRD) UNIV LELAND STANFORD JUNIOR.
PA (ORGA) ORGAMON NV.
PA
XX Hsueh AJW, Hsu SY, Liang S, Van Der Spek PJ;
PI WPI: 1999-591074/50.
XX N-PSDB: AA225346.
DR
XX
XX New G-protein coupled receptors, useful for identifying their own
PT ligands -
XX

PS Claim 2; Fig 4; 54pp; English.
 XX
 CC The present sequence represents the human G-protein coupled receptor
 CC having extracellular leucine rich repeat regions, designated LGR7 short
 CC form. The LGR4, LGR5 and LGR7 proteins are used to identify ligands for
 CC the receptor. The polypeptides and/or polynucleotides are also useful
 CC for homologous or related genes, producing compositions that modulate
 CC the expression or function of the receptors, gene therapy, mapping
 CC functional regions of the receptors, studying associated physiological
 CC pathways, in vivo prophylactic and therapeutic purposes, as immunogens
 CC for producing antibodies, and for identifying biologically active
 CC agents. The polypeptides contain a G-protein coupled seven
 CC transmembrane region and a leucine rich repeat extracellular domain.
 CC These regions capture and facilitate optimal orientation of its ligand.
 CC The proteins are also expressed in diverse tissues.
 XX
 XX Sequence 722 AA:

Query Match 100.0%; Score 3834; DB 20; Length 722;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 722; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTSGSVEFYLLIEGKFRSHGGGQDVKCSLGIFPCGNITKCLPOLLHCNGVDDCGNQADE 60
 DB 1 mtsqsvffyllifgkyfshgsgdvkcslygfpognltkclpqlhhngyddcgngaded 60
 QY 61 NCVVVLCQCSLPGLELDMKMPETSPVSNSNTAMSLQNLIRKLPDPCKNYHDLOKL 120
 DB 61 ncvvvlcqcsmlpgleldmkmptspvsnsntamslqnlirklpdpckknyhdlokl 120
 QY 121 DLQNNKTSISIVAFRGINSITLKLKYLNSHNRITFLKPGVFEDLHRLLEWLIIEDNHSRISP 180
 DB 121 dlqnnktsisivafrginsitlklkylnsnhrifltpgvfedlhrlewlilednhsrisp 180
 QY 181 PTFYGLNSLILVIMNNVTRLPDPLCOHMPRLHMLDEGNHINLRNTFTISCSMLTV 240
 DB 181 ptfyglnsllilvimmnvtrlpdkplcqhmprlhmldegnhlnlrntftiscsmly 240
 QY 241 LVMRKNINLHNTFAPLCKIDELDLSNKTIELPLIRFKDKELSGOLNNTYPIQIKQ 300
 DB 241 lvmrkninhntfaplcldeldlsnktielplirfkdkelsgolnntypiqikq 300
 QY 301 ANQFDYLVKLSLSLEGIEISNTIQRMFRPLMNLSHYFRKKFOYGYAPHYRSCKPTDG 360
 DB 301 anqfdylvklslslegieisnqrmfrplmnlshlyfkkfgygyapmyrscpkptdg 360
 QY 361 ISSLENLASTIQRVFWVWVSAYTCFENITVCRPIRSENNKLYAMSIISLCCADCLMG 420
 DB 361 isslenlasiqrvfwvwvsaytcfnitvcmprirsenklyamsiislccadclmg 420
 QY 421 IYLFVIGFPLKFGGEYKNAQQLMESTHCOLVSLAIIESTEVALLTFTLEKTYCIY 480
 DB 421 iylfvigfdlkgfgeynknaqlmesthcolvsglaaistevalltftlektyciy 480
 QY 481 YPRCVRPGKCRITVILILMITGFIATIPLSNKEFKNYGTNGVCFPLHSEDTESIG 540
 DB 481 yprcvrpgkcrityvillmitgfiatiplsnkefknygtngvcfplhsedestig 540
 QY 541 AQIYSVAIFGINNAARIIIVFSGSMFYSVHOSATRTETIRNKKEMILAKREFFIVF 600
 DB 541 aqiyvaifginnaariiivfsgsmfysvhosatrtetirnnkkmilakreffivf 600
 QY 601 TDALCWPIFVNFELSLQVEIRPCTISWVVFILPINSALNPILLYLTTRPEKEMHRE 660
 DB 601 tdcawpifvnfelslqveirpctiswvfvilpinsalnplillylttrpekemhrrf 660
 QY 661 WYNYRQRKSDSGQKTYAPSFITWEMVPLQEMPPELMKBDLFTYPCEMSLISOSTRLNS 720
 DB 661 wynyrqkmsdsgqkyapsfitwemvplqemppeilmkboldftyphemslisgstrlns 720
 QY 721 YS 722
 DB 721 YS 722

DB 721 ys 722

RESULT 2

ID AAY42170 standard; Protein; 757 AA.

AC AAY42170;

DT 20-DEC-1999 (first entry)

DE Human LGR7 long form protein sequence.

KW Human; LGR4; LGR5; LGR7; G-protein coupled receptor; gene therapy;
 KM extracellular leucine rich repeat region; mapping; identification.

OS Homo sapiens.

PN WO948921-A1.

PD 30-SEP-1999.

PF 25-MAR-1999; 99MO-US06573.

PR 26-MAR-1998; 98US-0079501.

PA (STRD) UNIV LELAND STANFORD JUNIOR.
 PA (ORGA) ORGANON NV.

PI Hsueh AJW, Hsu SY, Liang S, Van Der Spek PJ;

DR WPI; 1999-591074/50.

XX N-PDSB; AA225345.

PT New G-protein coupled receptors, useful for identifying their own
 PT ligands -

PS Claim 2; Fig 3; 54pp; English.

XX
 CC The present sequence represents the human G-protein coupled receptor
 CC having extracellular leucine rich repeat regions, designated LGR7 long
 CC form. The LGR4, LGR5 and LGR7 proteins are used to identify ligands for
 CC the receptor. The polypeptides and/or polynucleotides are also useful
 CC for homologous or related genes, producing compositions that modulate
 CC the expression or function of the receptors, gene therapy, mapping
 CC functional regions of the receptors, studying associated physiological
 CC pathways, in vivo prophylactic and therapeutic purposes, as immunogens
 CC for producing antibodies, and for identifying biologically active
 CC agents. The polypeptides contain a G-protein coupled seven
 CC transmembrane region and a leucine rich repeat extracellular domain.
 CC These regions capture and facilitate optimal orientation of its ligand.
 CC The proteins are also expressed in diverse tissues.

SQ Sequence 757 AA:

Query Match 96.2%; Score 3687.5; DB 20; Length 757;
 Best Local Similarity 93.0%; Pred. No. 0;
 Matches 707; Conservative 2; Mismatches 10; Indels 41; Gaps 3;

QY 1 MTSGSVEFYLLIEGKFRSHGGGQDVKCSLGIFPCGNITKCLPOLLHCNGVDDCGNQADE 60
 DB 1 mtsqsvffyllifgkyfshgsgdvkcslygfpognltkclpqlhhngyddcgngaded 60
 QY 61 NC-----VVVLCQCSLPGLELDMKMP- 82
 DB 61 ncgdingwsmqfdkyfasykmtsqypfeaecepclygsvpyvcloq--gledcdetn 117
 QY 83 FTSVPSSSVNTAMSLQNLIRKLPDCKFNKYHDLOKLQNNKITSISIVAFRGINSIT 142
 DB 118 ltravpsvsnvtamslqnlirklpdpckfnkynhdloqlqnnkitsisivafrginsit 177
 QY 143 KLYLSHNRITFLKPGVFEDLHRLLEWLIIEDNHSRISPTFYGLNSLILVIMNNVTRL 202

```
|||||
Db 178 klyshnrtiflpgvfedhrlweliieghnhsrispctfyglnsllllvlnmvltlrl 237
QY 203 PDKPLOCMPRLHMLDLEGNHINLNLNLFISCSNLTVLVNRKKNINHLNENFAPLQKL 262
Db 238 pdkplocgmprlhwldlegnhlnlrltfiscsnltvlvnrkknlnhlneentfapqlk 297
QY 263 DELDLGSKTENLPPLFKRLKELSOENLSTYNPORKIOANOFDVLVTKSLSEGTISN 322
Db 298 deldlgsnktenlpplfkrlkelsqnlstynporkiqanqfdylvtkslslegleisn 357
QY 323 IQQRMERPLNLNLSHIYKKFQYCGAPRHVSCKPNMDSISLENLASIIQRFVFWVYSA 382
Db 358 lqgrmfplnlnshlyfkktqyqyaprhvscokpntdgsislenllasllqrvfwvysa 417
QY 383 VTCEGNITVICMRRPYIRSENKLIAMSIISCCADCLMGITLVYIGFDLKFGEYKHAQ 442
Db 418 vtcfgnitfvcmrpyirsenklyamsislsccadclmgitylvigfdlkfrgeynkhaq 477
QY 443 LMWESFHCOLVGLSLALSTEVSVLLLFELTEKYICIVYPRCVRPGRKCTITVLLIWI 502
Db 478 lmweshcolvgsalstevsvlllfeltekycivypfrcvrgpkrctltvlllwi 537
QY 503 TGFIVAFIPLSNKEFEKNYGTNGVCPFLHSEDESIGAQIYSVAIFLGINLAFFITIVF 562
Db 538 tgfivafiplsnkefknnygtngvcfplhsedesigadqysvaiflglnlaaffitv 597
QY 563 SYGSMFYSVQSATITATEIRNOVKKEMLAKRPFYITFDALCMIPIFYVKSLSLOVEI 622
Db 598 sygsmfysvqsaltateirnvkkelakrffifvftdalcmipifvykflslslqvel 657
QY 623 PGRTSWVVFILPINSALNPILYTLFTRPFKEMIHREWYNRYORKMDSGOKTYPSF 682
Db 658 pgrtswvfvilpinalnpllytlftprfkemihrfwnygrksmdskgqktypsf 717
QY 683 IWEMMPLOEMPELMKPDLTFTYPCENSLISQSTRLSYS 722
Db 718 iwemmploempelmkpdltftypcenmsltsqstrlinsys 757

RESULT 3
AAE02498
ID AAE02498 standard; Protein; 396 AA.
XX
AC AAE02498;
XX
DE 10-AUG-2001 (first entry)
XX
DE Human CON222 G protein-coupled receptor protein.
XX
KW Human; G protein-coupled receptor; GPCR; CON222 protein; schizophrenia;
KW neuroleptic; nootropic; neuroprotective; bipolar disease; psychotropic;
KW neurological disorder; psychiatric disease; neurosis; anxiety; neuritis;
KW attention deficit hyperactivity disorder; neurosenia; senile dementia;
KW affective disorder; neuropathy; Alzheimer's disease; Parkinson's disease;
KW depression; migraine; genetic screening.
XX
OS Homo sapiens.
XX
FH Key
FH Domain
FT 42..65 /label= Transmembrane_domain_(1TW)
FT 66..78 /label= Intracellular_domain
FT /note= "First IC loop"
FT 79..103 /label= "First IC loop"
FT 104..124 /label= Transmembrane_domain_(2TW)
FT /label= Extracellular_domain
FT /note= "First EC loop"
FT 125..156 /label= Transmembrane_domain_(3TW)
FT 157..166 /label= Transmembrane_domain_(3TW)
FT Domain
```

```
FT /label= Intracellular_domain
FT /note= "Second IC loop"
FT 167..188 /label= Transmembrane_domain_(4TW)
FT 189..216 /label= Extracellular_domain
FT /label= "Second EC loop"
FT 217..241 /label= Transmembrane_domain_(5TW)
FT 242..267 /label= Intracellular_domain
FT /note= "Third IC loop"
FT 268..290 /label= Transmembrane_domain_(6TW)
FT 291..300 /label= Extracellular_domain
FT /note= "Third EC loop"
FT 301..320 /label= Transmembrane_domain_(7TW)
FT Domain
FT
FT
FT
XX WO200131014-A2.
XX
XX 03-MAY-2001.
XX
XX 27-OCT-2000; 2000WO-US29601.
XX
XX 27-OCT-1999; 99US-0427653.
XX 27-OCT-1999; 99US-0427859.
XX 27-OCT-1999; 99US-0428020.
XX 27-OCT-1999; 99US-0428114.
XX 28-OCT-1999; 99US-0429517.
XX 28-OCT-1999; 99US-0429555.
XX 28-OCT-1999; 99US-0429676.
XX 28-OCT-1999; 99US-0429695.
XX 03-DEC-1999; 99US-0454399.
XX 12-JAN-2000; 2000US-0481794.
XX
XX (PHAA ) PHARMACIA & UPJOHN CO.
XX
XX Vogel G, Wood LS, Merchant K;
XX
XX WPI; 2001-328653/34.
XX
XX N-PSDB; AAD06507.
XX
XX
XX Seven transmembrane receptor polypeptides and polynucleotides, useful
XX for treating neurological or psychiatric disorders, e.g. schizophrenia,
XX as well as for identifying compounds useful for treating schizophrenia
XX
XX
XX Claim 1; Page 15-16; 215pp; English.
XX
XX
XX The invention relates to human G protein-coupled receptor (GPCR) and
XX their corresponding DNA molecules. GPCR is also referred as seven
XX transmembrane receptor. G protein-coupled receptor protein is useful for
XX treating neurological disorder, particularly schizophrenia. GPCR protein
XX is also useful for identifying compounds useful for treating other
XX schizophrenia. These compounds are also useful for treating other
XX disease, affective and psychiatric diseases, e.g. depression, anxiety, bipolar
XX attention deficit disorder, epilepsy, neuritis, neurosenia, neuropathy,
XX neurosis, Alzheimer's disease, Parkinson's disease, migraine and senile
XX dementia. The invention also provides genetic screening procedures that
XX entail analysing a person's genome with respect to GPCR. The vectors are
XX useful for the recombinant production of the GPCR's. The present sequence
XX is human CON222 G protein-coupled receptor (GPCR) protein.
XX
XX
XX Sequence 396 AA;
```

```
Query Match 54.5%; Score 2091; DB 22; Length 396;
Best Local Similarity 99.7%; Pred. No. 7e-195;
Matches 395; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

QY	327	MFRLMLNSHIYFRKKPOYCGYAPHVRSCKRNTDGSISEMLLASIORPVMVWSAVCF	386
Db	1	mftrplvnlshyfkfkygcyaphvrsckpntdglslenllasllqtrfvwvssavtcf	60
QY	387	GNIVICMRPIYRSEKNTLYAMSIIISLCCADCLMGIFYLFYVGFDFLEFRGEYNKHAQJAME	446
Db	61	gnlfvycmryirsenklyamsiisliccadclmgifylfvfgfdlfrgeynkhaqlame	120
QY	447	STHCQVGSALISLEVSVALTFLELTLEKYICIAVPEPRCVRPCKRTIYVLLIMITGEI	506
Db	121	sthcqlvglsalislstevsvalttfletlekyiciahyprevrpkrcrtitvllilmitgfi	180
QY	507	VAFIPLSKKEFKNYCYTCNGYCPFLHSEDPHESTGCAQIYSAIFLGINLAAFIITVSYGS	566
Db	181	vafiplskkefknykytngyvcplhshedcsjgaqiyvsaiiflglnlaaflilivisygs	240
QY	567	MEYSHOSATATELRNOVKKEMILAKREFFIYTDLMCIPIFVAFESLLOVEIIPGTI	626
Db	241	mfsyvhgsatateletrngyvkemilakrffiffiytdlmcwiplfvvflsljqlveipgtl	300
QY	627	TSWVVFILPIINSALNPILLYLTTRPRKEMIHREFWNYRORKSMDSKGOKTVAPSPFIWE	686
Db	301	tswvffilpfnalsalnpiljlylttrpfkemmhrfwyrgqksmdskgktyapsfiawe	360
QY	687	MMPIQMBEPELAKRDPYTPCEMSLIQSOTRFLNYS	722
Db	361	mmpiqempepelmkpdlfitypcemslisqstrflnys	396

RESULT 4
 ID AAY57286 standard: Protein; 396 AA.
 AC AAY57286;
 DT 05-JUN-2000 (first entry)
 DE Human GPCR protein (HGPRP) sequence (clone ID 2488822).
 KW Human; G protein coupled protein receptor; HGPRP; cell proliferation;
 KW neurological; immune disorder; cytostatic; anti-arteriosclerotic;
 KW anti-atherosclerotic; hepatotropic; antiinflammatory; virocidic; leukemia;
 KW immunomodulatory; anemia; asthma; gastrointestinal; anti-epileptic;
 KW anti-Alzheimer's; anti-Parkinsonian; gene therapy.
 OS Homo sapiens.
 PN WO200015793-A2.
 PD 23-MAR-2000.
 PF 17-SEP-1999; 99WO-US20958.
 PR 17-SEP-1998; 98US-0156513.
 PA (INCY-) INCYTE PHARM INC.
 PI Bandman O, Lal P, Tang YT, Corley NC, Guegler KU, Gorgone GA;
 PI Baughn MR;
 DR MPI: 2000-271432/23.
 DR N-PSDB; AAZ90524.
 PT Human G protein coupled protein receptor peptides useful for the
 PT prevention, diagnosis and treatment of cell proliferative, neurological
 and immune disorders -
 PS Claim 1; Page 62-63; 71pp: English.
 CC The invention provides human G protein coupled protein receptor (HGPRP)
 CC polypeptides and polynucleotides encoding them. The polypeptides can be
 CC produced by standard recombinant methodology. The polynucleotides and
 CC polypeptides may be used in the prevention, treatment and diagnosis of

CC diseases associated with their inappropriate expression. Diseases that
CC can be treated are cell proliferative disorders (e.g. arteriosclerosis,
CC atherosclerosis and hepatitis), cancers (e.g. leukemia, melanomas and
CC adenocarcinoma), immune disorders (e.g. anemia, asthma and Crohn's
CC disease) and neurological disorders (e.g. epilepsy, Alzheimer's disease
CC and Parkinson's disease). The anti-HGPR antibodies may also be used as
CC diagnostic agents for detecting the presence of HGPR polypeptides in
CC samples (e.g. by enzyme linked immunosorbant assay (ELISA)). Sequences
CC AA57283-288 represent the HGPR polypeptides.

	Query Match	54.2%	Score 2079;	DB 21;	Length 396;	
	Best Local Similarity	99.2%;	Pred. No. 1e-193,			
	Matches 393;	Conservative 1;	Mismatches 2;	Indels 0;	Gaps 0;	
OY	327 MRRPLMNLSHIFYKFKFEQCGVAPHVRSCCKPNNDGISLSENLASTIQRVFVWWSAVTCF	386				
Db	1 MRRPLVNLSHIYFKFTFYGQGYAPHRSCCKPNTDGLSLLENLAISIIQRFVWVSATCF	60				
OY	387 GNIFVICMRRPIRSENKIYAMSIIISLCADCIMGIYLFYIGGFDLKFGEYKKHQLMNE	446				
Db	61 gnlflvcmrpyrrsenklyamsisiscadclmgilylfviyigfdlktfgeynkhaqlme	120				
OY	447 SHHCQLVGSLALTSREVSALLTFLLTEKYICIVPEPRCVRPGKCRTIVLLIWITGFI	506				
Db	121 sthcqrlvgslalstevsvllltfltlekyicivpfrcvrpgkcrtllvlllwitgfi	180				
OY	507 VAFIPISNKEEFKNYGTGVGCPELHSEBTESIGAQISVAIFLGINLAFTIIVFSYS	566				
Db	181 valfpisnkeeffknypapgvvcfphsdteeisgaqiyvalfglnlaaffilivfsys	240				
OY	567 MFYSVHQSAIYATEIRNOVKKEKITAKRFFIYFPDIALCWIPFYVKKFSLLOVEHPGT	626				
Db	241 mfyssvhqsaitateirngvkemilaakrfiflvftdalcwiplfvkkfslsqvelipgtl	300				
OY	627 TSWVVVFILIPINSALNPILIYTLTRPFKEMIHREFWYNRQORSMDSGOKTAPSEFWE	686				
Db	301 tswvvvfllpinsaenlpillytltrpfkemlnrfwyny-qrkmsdskgqktyapsefiwe	360				
OY	687 MMFLQEMPEELMKPDIFTYPCEMSLISOSTRLNSYS	722				
Db	361 mwplgempeelmkpdlftyppcemslisgstrlnsys	386				
RESULT	5					
AAU04370	AAU04370 standard; Protein; 355 AA.					
XX	AAU04370;					
AC	AAU04370;					
XX	23-OCT-2001 (first entry)					
DT						
XX	Human G-protein coupled receptor, hRUP16.					
DE						
XX	Human: G-protein coupled receptor; GPCR; hRUP16; agonist;					
KM	Inverse agonist; lung cancer.					
XX	Homo sapiens.					
OS						
XX	WO200136471-A2.					
PN						
XX	25-MAY-2001.					
PD						
XX	16-NOV-2000; 2000MO-US31509.					
PF						
XX	17-NOV-1999; 99US-0166088.					
PR	17-NOV-1999; 99US-0166099.					
PR	17-NOV-1999; 99US-0166369.					
PR	23-DEC-1999; 99US-0171900.					
PR	23-DEC-1999; 99US-0171901.					
PR	23-DEC-1999; 99US-0171902.					

PR 11-FEB-2000; 2000US-0181749.
 PR 14-MAR-2000; 2000US-0189258.
 PR 14-MAR-2000; 2000US-0189259.
 PR 10-APR-2000; 2000US-0195898.
 PR 10-APR-2000; 2000US-0195899.
 PR 10-APR-2000; 2000US-0196078.
 PR 28-APR-2000; 2000US-0200419.
 PR 12-MAY-2000; 2000US-0203630.
 PR 12-JUN-2000; 2000US-0210741.
 PR 12-JUN-2000; 2000US-0210982.
 PR 21-AUG-2000; 2000US-0226760.
 PR 26-SEP-2000; 2000US-0235418.
 PR 26-SEP-2000; 2000US-0235779.
 PR 20-OCT-2000; 2000US-0242332.
 PR 20-OCT-2000; 2000US-0242343.
 XX
 PA (AREN-) ARENA PHARM INC.
 XX
 FI Chen R, Dang HT, Lowitz KP;
 FI
 DR WPI; 2001-355616/37.
 DR N-PSDB; AAS07943.
 XX
 PT Endogenous and non-endogenous versions of human G-protein coupled
 PT receptors for direct identification of candidate compounds as agonists,
 PT inverse agonists or partial agonists for use as therapeutic agents -
 PS
 PS Claim 33; Page 105-106; 160pp; English.
 CC The sequence represents a human G-protein coupled receptor (GPCR),
 CC hRPR16. The endogenous and non-endogenous, constitutively activated
 CC versions of human G-protein coupled receptors (GPCR), are useful for
 CC direct identification of candidate compounds as receptor agonists,
 CC inverse agonists or partial agonists having applicability as therapeutic
 CC agents for treating diseases related to GPCR, e.g. lung cancer.
 CC Non-endogenous version of human GPCRs are also utilized in research
 CC settings and in vitro and in vivo system, incorporating GPCRs can be
 CC utilised to elucidate and understand the roles these receptors
 CC play in the human condition, both normal and diseased.
 XX
 SQ Sequence 355 AA;
 SQ
 Query Match 28.2%; Score 1082.5; DB 22; Length 355;
 Best Local Similarity 59.8%; Pred. No. 1.le-96;
 Matches 198; Conservative 67; Mismatches 65; Indels 1; Gaps 1
 QY 356 PNTDGISSLENIILASIIQREVEVWVSAYTCFENIFVICRPIRSENKLYAMSTISLCCA 415
 Db 2 pltdgiSSFedllannllirllfwvafllcfnllfygmrsfkkaentthamsikllcca 61
 QY 416 DCLMIVTIFVIGGPELKRGRGEYKNRKAQOLMESTHOQVSLAILSEVSVLLTFTLEK 475
 Db 62 dclmivgtifvgyifdkrygqyqkallmmsvqcrllngflamlstevsvllltyltlek 121
 QY 476 YICIYVPPRCVAPGRCRITTVLILMINGFIYAFLPSNKEFFKNYGRNGVCPHLSDD 535
 Db 122 flvtyfppsnrlpygrkrqcvsltlclmmgfflaavlpfnmkdyfgnfykngvcfplpydq 181
 QY 536 TESTGAQIVSAITPLGINLAFTIIVFSYSGSMFYSVHQSATATETIRNOVKKEMILAKRF 595
 Db 182 tedtsgskyslgiflvgvnlalafllivfsytlmfcsiqktaqlctvevncfgrvavannf 241
 QY 596 FEIVTDLQWPIRVVNFSLLOVEIPEITISWVFIPLINSANLPILYLTTPPEK 655
 Db 242 fivivsdaiqlwpiwvnlslflveipdltmcsvvliflvpnsalnpllyltlnfkd 301
 QY 656 MHRFWYNYRQRKSDMSKQKTYAPSFIVWE 686
 Db 302 ktkqlhkh-grksifklkkkslsctsiwle 331

PR	08-SEP-2000	2000US-0231414	2000US-0231414
PR	08-SEP-2000	2000US-0232080	2000US-0232080
PR	08-SEP-2000	2000US-0232081	2000US-0232081
PR	12-SEP-2000	2000US-0231968	2000US-0231968
PR	14-SEP-2000	2000US-0232297	2000US-0232297
PR	14-SEP-2000	2000US-0232298	2000US-0232298
PR	14-SEP-2000	2000US-0232299	2000US-0232299
PR	14-SEP-2000	2000US-0232400	2000US-0232400
PR	14-SEP-2000	2000US-0232401	2000US-0232401
PR	14-SEP-2000	2000US-0233063	2000US-0233063
PR	14-SEP-2000	2000US-0233064	2000US-0233064
PR	14-SEP-2000	2000US-0233065	2000US-0233065
PR	21-SEP-2000	2000US-0234423	2000US-0234423
PR	21-SEP-2000	2000US-0234424	2000US-0234424
PR	25-SEP-2000	2000US-0234997	2000US-0234997
PR	25-SEP-2000	2000US-0234998	2000US-0234998
PR	25-SEP-2000	2000US-0235684	2000US-0235684
PR	26-SEP-2000	2000US-0235684	2000US-0235684
PR	27-SEP-2000	2000US-0235834	2000US-0235834
PR	27-SEP-2000	2000US-0235836	2000US-0235836
PR	29-SEP-2000	2000US-0236327	2000US-0236327
PR	29-SEP-2000	2000US-0236327	2000US-0236327
PR	29-SEP-2000	2000US-0236568	2000US-0236568
PR	29-SEP-2000	2000US-0236569	2000US-0236569
PR	29-SEP-2000	2000US-0236570	2000US-0236570
PR	02-OCT-2000	2000US-0237032	2000US-0237032
PR	02-OCT-2000	2000US-0237037	2000US-0237037
PR	02-OCT-2000	2000US-0237038	2000US-0237038
PR	02-OCT-2000	2000US-0237039	2000US-0237039
PR	02-OCT-2000	2000US-0237040	2000US-0237040
PR	02-OCT-2000	2000US-0237045	2000US-0237045
PR	13-OCT-2000	2000US-0239935	2000US-0239935
PR	13-OCT-2000	2000US-0239937	2000US-0239937
PR	20-OCT-2000	2000US-0240960	2000US-0240960
PR	20-OCT-2000	2000US-0241221	2000US-0241221
PR	20-OCT-2000	2000US-0241785	2000US-0241785
PR	20-OCT-2000	2000US-0241786	2000US-0241786
PR	20-OCT-2000	2000US-0241787	2000US-0241787
PR	20-OCT-2000	2000US-0241808	2000US-0241808
PR	20-OCT-2000	2000US-0241809	2000US-0241809
PR	20-OCT-2000	2000US-0244617	2000US-0244617
PR	01-NOV-2000	2000US-0244617	2000US-0244617
PR	08-NOV-2000	2000US-0246474	2000US-0246474
PR	08-NOV-2000	2000US-0246475	2000US-0246475
PR	08-NOV-2000	2000US-0246526	2000US-0246526
PR	08-NOV-2000	2000US-0246527	2000US-0246527
PR	08-NOV-2000	2000US-0246528	2000US-0246528
PR	08-NOV-2000	2000US-0246532	2000US-0246532
PR	08-NOV-2000	2000US-0246533	2000US-0246533
PR	08-NOV-2000	2000US-0246524	2000US-0246524
PR	08-NOV-2000	2000US-0246525	2000US-0246525
PR	08-NOV-2000	2000US-0246526	2000US-0246526
PR	08-NOV-2000	2000US-0246511	2000US-0246511
PR	08-NOV-2000	2000US-0246613	2000US-0246613
PR	17-NOV-2000	2000US-0248207	2000US-0248207
PR	17-NOV-2000	2000US-0248208	2000US-0248208
PR	17-NOV-2000	2000US-0248210	2000US-0248210
PR	17-NOV-2000	2000US-0248211	2000US-0248211
PR	17-NOV-2000	2000US-0248212	2000US-0248212
PR	17-NOV-2000	2000US-0248213	2000US-0248213
PR	17-NOV-2000	2000US-0248214	2000US-0248214
PR	17-NOV-2000	2000US-0248215	2000US-0248215
PR	17-NOV-2000	2000US-0248216	2000US-0248216
PR	17-NOV-2000	2000US-0248217	2000US-0248217
PR	17-NOV-2000	2000US-0248218	2000US-0248218
PR	17-NOV-2000	2000US-0248244	2000US-0248244
PR	17-NOV-2000	2000US-0248245	2000US-0248245
PR	17-NOV-2000	2000US-0248246	2000US

17-NOV-2000; 2000US-0249309.
PR
PR 17-NOV-2000; 2000US-0249309.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250330.
PR 05-DEC-2000; 2000US-0251039.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251866.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCT INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
PI
DR MPI: 2001-465573/.
DR N-PSTDB: AAI99557.
XX
PT Isolated digestive system associated polypeptide for treating,
PT preventing and/or prognosing disorders related to the digestive system
PT including digestive system cancers and also for testing and detection
PT e.g. diagnosis -
XX
Claim 11; SEQ ID NO 69; 509pp + Sequence Listing; English.

The invention relates to novel genes (AAI99548-AAI99604) and proteins (AAI99936-AAI99984) useful for preventing, treating or ameliorating medical conditions e.g. by protein or gene therapy. The genes are isolated from a range of human tissues disclosed in the specification. The nucleic acids, proteins, antibodies and (ant)agonists are useful in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and ovarian cancer and other cancers of the adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver, lung, or urogenital;

(b) immune disorders e.g. Addison's disease, allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c) cardiovascular disorders such as myocardial ischaemias; (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f) infectious diseases such as viral, bacterial, fungal and parasitic infections.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

Sequence 188 AA:

Query Match	24.7%	Score 948	DB 22	Length 188
Best Local Similarity	100.0%	Pred. No. 5,7e-84		
Matches 182	Conservative 0	Mismatches 0	Indels 0	Gaps 0
OY	541	AOIYSVAIFLGINLAAFIIVSYGSMFYSVQASITATAIRNQYKKEMILAKREFEIVF	600	
Db	7	aqiyvalflglnlaaflifivsgysmfsvngsataleirngykkemilakrfflff	66	
OY	601	TDALCWIPLEFVVKELSLQVEIPGTTISWVFIILPINSALNPILYTLITRPFKMIREF	660	
Db	67	tdalcwlpflfvkflslsqveipgtswwvflilpnsalnpllytlitripfkemlrff	126	
OY	661	WYNVYORKSMDSGOKTYAPSPFIWEMKMPLOEMPELKKPDLETFYPCESLSISQSTRINS	720	
Db	127	wynvyrksmsdsgoktyapsfiiwemmploempelnkpdflfypcemslisqstrins	186	
OY	721	YS 722		
Db	187	ys 188		

RESULT 7
AAM9972 ID AAM9972 standard; Protein: 176 AA.
XX AAM9972;
AC
XX
DT 04-JAN-2002 (first entry)
XX
DE Human expressed polypeptide SEQ ID NO 96.
XX
KW Human; nootropic; neuroprotective; cytosolic; dermatological; virucide;
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnery;
KW antiparkinsonian; antisticking; antianemic; antiarthritic; cancer;
KW antineumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW antiallergic; antidiabetic; antilicer; anticonvulsant; antifungal;
KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine.
XX
OS Homo sapiens.
XX
PN W0200155387-A1.
XX
PD 02-AUG-2001.
XX
PE 17-JAN-2001; 2001WO-US01310.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.

PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233065.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241121.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.

PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251988.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-465573/50.
DR N-PSDB; AAI99584.
XX
XX Isolated digestive system associated polypeptide for treating,
PT preventing and/or prognosing disorders related to the digestive system
PT including digestive system cancers and also for testing and detection
PT e.g. diagnosis -
XX
PS Claim 11; SEQ ID NO 96; 509pp + Sequence listing; English.
XX
XX The invention relates to novel genes (AAI99548-AAI99604) and proteins
CC (AAM9936-AA99984) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 176 AA;
Query Match 23.8%; Score 912; DB 22; Length 176;
Best Local Similarity 99.4%; Pred. No. 1,7e-80;
Matches 175; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 541 AQTSAVFLEINLAFFIIFVSGMYSVHOSATITATEIRNOVKKEMIAKREFEIV 600
Db 1 aqlysvaffinlaaffilivfsgmlysvgsaltateirngkkmllakrtiflvf 60
QY 601 TDAICWPIFEVVKFLSLQVEIPGTTISWVIFILPINSALNPILYLTTPRFKEMIRF 660
Db 61 tdaicwplfvkflslqlveipgltiswvifilpnsalnpllyltlttprfkemirf 120
QY 661 WYNTRKRSKMSDKQKTYAPSFIVEMKPILOEMPEELMKPDLFFYPCMSLISQST 716
Db 121 wyntrqksmdskgktyapsfivemwplqempelmkpldlffypcmslissgt 176

XX
AC ABB04062;
XX
DT 08-JAN-2002 (first entry)
XX
DE Human musculoskeletal system related polypeptide SEQ ID NO 2009.
XX
KW Cytostatic; immunosuppressive; neutropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiact; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein;
KW musculoskeletal system.
XX
OS Homo sapiens.
XX
PN WO20015367-A1.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01338.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229345.
PR 01-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 05-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.

DT 10-JAN-2002 (first entry)
XX Human CDNA SEQ ID NO: 557.
XX
KW Human; gene therapy; neural disorder; immune system disorder;
muscular disorder; reproductive disorder; gastrointestinal disorder;
pulmonary disorder; cardiovascular disorder; renal disorder;
proliferative disorder; inflammation.
XX Homo sapiens.
XX WO200154474-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01349.
XX
PR 31-JAN-2000; 2000US-179065P.
PR 04-FEB-2000; 2000US-180628P.
PR 24-FEB-2000; 2000US-184664P.
PR 02-MAR-2000; 2000US-186350P.
PR 16-MAR-2000; 2000US-189874P.
PR 17-MAR-2000; 2000US-190076P.
PR 18-APR-2000; 2000US-198123P.
PR 19-MAY-2000; 2000US-205515P.
PR 07-JUN-2000; 2000US-209467P.
PR 28-JUN-2000; 2000US-214886P.
PR 30-JUN-2000; 2000US-215135P.
PR 07-JUL-2000; 2000US-216647P.
PR 07-JUL-2000; 2000US-216880P.
PR 11-JUL-2000; 2000US-217487P.
PR 11-JUL-2000; 2000US-217496P.
PR 14-JUL-2000; 2000US-218290P.
PR 26-JUL-2000; 2000US-220963P.
PR 26-JUL-2000; 2000US-220964P.
PR 14-AUG-2000; 2000US-224518P.
PR 14-AUG-2000; 2000US-224519P.
PR 14-AUG-2000; 2000US-225213P.
PR 14-AUG-2000; 2000US-225214P.
PR 14-AUG-2000; 2000US-225267P.
PR 14-AUG-2000; 2000US-225268P.
PR 14-AUG-2000; 2000US-225270P.
PR 14-AUG-2000; 2000US-225447P.
PR 14-AUG-2000; 2000US-225757P.
PR 14-AUG-2000; 2000US-225758P.
PR 14-AUG-2000; 2000US-225759P.
PR 18-AUG-2000; 2000US-226279P.
PR 22-AUG-2000; 2000US-226811P.
PR 22-AUG-2000; 2000US-226868P.
PR 22-AUG-2000; 2000US-227182P.
PR 23-AUG-2000; 2000US-227009P.
PR 30-AUG-2000; 2000US-228924P.
PR 01-SEP-2000; 2000US-229287P.
PR 01-SEP-2000; 2000US-229343P.
PR 01-SEP-2000; 2000US-229344P.
PR 01-SEP-2000; 2000US-229345P.
PR 05-SEP-2000; 2000US-229509P.
PR 05-SEP-2000; 2000US-229513P.
PR 06-SEP-2000; 2000US-230437P.
PR 06-SEP-2000; 2000US-230438P.
PR 08-SEP-2000; 2000US-231242P.
PR 08-SEP-2000; 2000US-231243P.
PR 08-SEP-2000; 2000US-231244P.
PR 08-SEP-2000; 2000US-231413P.
PR 08-SEP-2000; 2000US-231414P.
PR 08-SEP-2000; 2000US-232089P.
PR 08-SEP-2000; 2000US-232081P.
PR 12-SEP-2000; 2000US-231968P.
PR 14-SEP-2000; 2000US-232397P.
PR 14-SEP-2000; 2000US-232398P.
PR 14-SEP-2000; 2000US-232399P.
PR 14-SEP-2000; 2000US-232400P.

PR 14-SEP-2000; 2000US-232401P.
PR 14-SEP-2000; 2000US-233063P.
PR 14-SEP-2000; 2000US-233064P.
PR 21-SEP-2000; 2000US-233065P.
PR 21-SEP-2000; 2000US-234223P.
PR 21-SEP-2000; 2000US-234274P.
PR 25-SEP-2000; 2000US-234997P.
PR 25-SEP-2000; 2000US-234998P.
PR 26-SEP-2000; 2000US-235844P.
PR 27-SEP-2000; 2000US-235834P.
PR 27-SEP-2000; 2000US-235836P.
PR 29-SEP-2000; 2000US-236327P.
PR 29-SEP-2000; 2000US-236367P.
PR 29-SEP-2000; 2000US-236368P.
PR 29-SEP-2000; 2000US-236369P.
PR 29-SEP-2000; 2000US-236370P.
PR 02-OCT-2000; 2000US-236802P.
PR 02-OCT-2000; 2000US-237037P.
PR 02-OCT-2000; 2000US-237038P.
PR 02-OCT-2000; 2000US-237039P.
PR 02-OCT-2000; 2000US-237040P.
PR 13-OCT-2000; 2000US-239335P.
PR 13-OCT-2000; 2000US-239337P.
PR 20-OCT-2000; 2000US-240960P.
PR 20-OCT-2000; 2000US-241221P.
PR 20-OCT-2000; 2000US-241785P.
PR 20-OCT-2000; 2000US-241786P.
PR 20-OCT-2000; 2000US-241787P.
PR 20-OCT-2000; 2000US-241808P.
PR 20-OCT-2000; 2000US-241809P.
PR 20-OCT-2000; 2000US-241826P.
PR 01-NOV-2000; 2000US-244617P.
PR 08-NOV-2000; 2000US-246474P.
PR 08-NOV-2000; 2000US-246475P.
PR 08-NOV-2000; 2000US-246476P.
PR 08-NOV-2000; 2000US-246477P.
PR 08-NOV-2000; 2000US-246478P.
PR 08-NOV-2000; 2000US-246523P.
PR 08-NOV-2000; 2000US-246524P.
PR 08-NOV-2000; 2000US-246525P.
PR 08-NOV-2000; 2000US-246526P.
PR 08-NOV-2000; 2000US-246527P.
PR 08-NOV-2000; 2000US-246528P.
PR 08-NOV-2000; 2000US-246532P.
PR 08-NOV-2000; 2000US-246609P.
PR 08-NOV-2000; 2000US-246610P.
PR 08-NOV-2000; 2000US-246611P.
PR 08-NOV-2000; 2000US-246613P.
PR 17-NOV-2000; 2000US-249207P.
PR 17-NOV-2000; 2000US-249208P.
PR 17-NOV-2000; 2000US-249209P.
PR 17-NOV-2000; 2000US-249210P.
PR 17-NOV-2000; 2000US-249211P.
PR 17-NOV-2000; 2000US-249212P.
PR 17-NOV-2000; 2000US-249213P.
PR 17-NOV-2000; 2000US-249214P.
PR 17-NOV-2000; 2000US-249215P.
PR 17-NOV-2000; 2000US-249216P.
PR 17-NOV-2000; 2000US-249217P.
PR 17-NOV-2000; 2000US-249218P.
PR 17-NOV-2000; 2000US-249244P.
PR 17-NOV-2000; 2000US-249245P.
PR 17-NOV-2000; 2000US-249246P.
PR 17-NOV-2000; 2000US-249256P.
PR 17-NOV-2000; 2000US-249257P.
PR 17-NOV-2000; 2000US-249297P.
PR 17-NOV-2000; 2000US-249300P.
PR 01-DEC-2000; 2000US-250160P.
PR 01-DEC-2000; 2000US-250391P.
PR 05-DEC-2000; 2000US-251030P.
PR 05-DEC-2000; 2000US-251988P.
PR 05-DEC-2000; 2000US-256719P.
PR 06-DEC-2000; 2000US-251479P.

PR 08-DEC-2000; 2000US-251856P.
PR 08-DEC-2000; 2000US-251868P.
PR 08-DEC-2000; 2000US-251869P.
PR 08-DEC-2000; 2000US-251889P.
PR 08-DEC-2000; 2000US-251906P.
PR 11-DEC-2000; 2000US-254097P.
PR 05-JAN-2001; 2001US-259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
PI
XX WPI; 2001-476161/51.
DR N-PSDB; ABA06471.
XX
XX
PT Isolated nucleic acid molecule encoding an inflammation-associated
PT polypeptide is used in preventing, treating or ameliorating a medical
PT condition .
XX
XX
PS Claim 11; SEQ ID NO: 557; 859pp + Sequence Listing; English.
XX
XX
CC The present invention provides human cDNAs, proteins and related genomic
CC DNAs. These can be used in the treatment of neural, immune system,
CC muscular, reproductive, gastrointestinal, pulmonary, cardiovascular,
CC renal and proliferative disorders and inflammation. The present sequence
CC is a protein of the invention.
XX
SQ Sequence 176 AA;

Query Match 23.8%; Score 912; DB 22; Length 176;
Best Local Similarity 99.4%; Pred. No. 1.7e-80;
Matches 175; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 541 AOIYVAIFLGINLAFTIIVFSGSMFYSHQSAITATEIRNOYKKEMILAKREFEIVF 600
DB 1 aqiyvaiflginlaaffifvsgsmfysvhsaiteirngkkemilakrffeflvf 60
OY 601 TDALCWIPFVYVKFLSLQVEIPGITTSWVVFILPINSALNPILYTLTRPFKEIHNF 660
DB 61 tcalcwipflfvkvkflslqveipgitstswvvlflpinalnplilytltrpfkemihrf 120
OY 661 WYNYRQRKSKMSDGGOTYAPSEFIWEMNPLOEMPPELMPDLFTYPCSEMSLSQST 716
DB 121 wynyrgqrksmdsggqktyapsfiwemnpdqempehmpdlftypcemslisqst 176
RESULT 10
AAU18108
ID AAU18108 standard; Protein: 176 AA.
XX
AC AAU18108;
XX
DT 21-NOV-2001 (first entry)
XX
XX Novel human uterine motility-association polypeptide #15.
DE
XX Human; uterine motility-association disorder; uterus; pregnancy;
KW labour; menstrual cycle; gene therapy.
XX
OS Homo sapiens.
XX
PM WO200155201-A1.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01317.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198122.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216680.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227182.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0228287.
PR 01-SEP-2000; 2000US-0228343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 06-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235835.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.

PR	13-OCT-2000;	2000US-0239937.	PR
PR	20-OCT-2000;	2000US-0240960.	PR
PR	20-OCT-2000;	2000US-0241221.	PR
PR	20-OCT-2000;	2000US-0241785.	PR
PR	20-OCT-2000;	2000US-0241786.	PR
PR	20-OCT-2000;	2000US-0241787.	PR
PR	20-OCT-2000;	2000US-0241808.	PR
PR	20-OCT-2000;	2000US-0241809.	PR
PR	20-OCT-2000;	2000US-0241826.	PR
PR	01-NOV-2000;	2000US-0244617.	PR
PR	08-NOV-2000;	2000US-0246474.	PR
PR	08-NOV-2000;	2000US-0246475.	PR
PR	08-NOV-2000;	2000US-0246476.	PR
PR	08-NOV-2000;	2000US-0246478.	PR
PR	08-NOV-2000;	2000US-0246523.	PR
PR	08-NOV-2000;	2000US-0246524.	PR
PR	08-NOV-2000;	2000US-0246525.	PR
PR	08-NOV-2000;	2000US-0246526.	PR
PR	08-NOV-2000;	2000US-0246527.	PR
PR	08-NOV-2000;	2000US-0246528.	PR
PR	08-NOV-2000;	2000US-0246532.	PR
PR	08-NOV-2000;	2000US-0246509.	PR
PR	08-NOV-2000;	2000US-0246610.	PR
PR	08-NOV-2000;	2000US-0246611.	PR
PR	08-NOV-2000;	2000US-0246613.	PR
PR	17-NOV-2000;	2000US-0249207.	PR
PR	17-NOV-2000;	2000US-0249208.	PR
PR	17-NOV-2000;	2000US-0249209.	PR
PR	17-NOV-2000;	2000US-0249210.	PR
PR	17-NOV-2000;	2000US-0249211.	PR
PR	17-NOV-2000;	2000US-0249212.	PR
PR	17-NOV-2000;	2000US-0249213.	PR
PR	17-NOV-2000;	2000US-0249214.	PR
PR	17-NOV-2000;	2000US-0249215.	PR
PR	17-NOV-2000;	2000US-0249216.	PR
PR	17-NOV-2000;	2000US-0249217.	PR
PR	17-NOV-2000;	2000US-0249218.	PR
PR	17-NOV-2000;	2000US-0249224.	PR
PR	17-NOV-2000;	2000US-0249245.	PR
PR	17-NOV-2000;	2000US-0249264.	PR
PR	17-NOV-2000;	2000US-0249265.	PR
PR	17-NOV-2000;	2000US-0249265.	PR
PR	17-NOV-2000;	2000US-0249297.	PR
PR	17-NOV-2000;	2000US-0249299.	PR
PR	17-NOV-2000;	2000US-0249300.	PR
PR	01-DEC-2000;	2000US-0250160.	PR
PR	01-DEC-2000;	2000US-0250391.	PR
PR	05-DEC-2000;	2000US-0251030.	PR
PR	05-DEC-2000;	2000US-0251968.	PR
PR	05-DEC-2000;	2000US-0256719.	PR
PR	06-DEC-2000;	2000US-0251479.	PR
PR	08-DEC-2000;	2000US-0251856.	PR
PR	08-DEC-2000;	2000US-0251868.	PR
PR	08-DEC-2000;	2000US-0251869.	PR
PR	08-DEC-2000;	2000US-0251989.	PR
PR	08-DEC-2000;	2000US-0251990.	PR
PR	11-DEC-2000;	2000US-0254097.	PR
PR	05-JAN-2001;	2001US-0259678.	PR

PA	(HUMA-) HUMAN GENOME SCI INC.
XX	
PI	Rosen CA, Barash SC, Ruben SM;
XX	
DR	WPI; 2001-488777/53.
DR	N-PSDE; AAS28950.
XX	
PT	Isolated polypeptide and nucleic acid molecules for treating,
PT	preventing and/or prognosing disorders related to uterine motility
PT	e.g. disorders associated with pregnancy and the menstrual cycle -
XX	
PS	Claim 11; SEQ ID NO 84; 524pp; English.
XX	
CC	The present invention relates to the isolation of novel human

uterine motility-association polypeptides, and cDNA (AA528936-AA528994) and genomic sequences encoding for these polypeptides. The sequences of the invention are useful in the diagnosis, treatment, prevention and/or prognosis of diseases associated with uterine motility such as pregnancy and labour, and menstrual disorders. The polynucleotide sequences of the invention are also useful in gene therapy. AA018094-AA018152 represent novel human uterine motility-association polypeptides. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

	Query Match	23.8%	Score 912	DB 22	Length 176
	Best Local Similarity	99.4%	Pred. No. 1.7e-80		
	Matches 175	Conservative	0	Matches 1	Indels 0
				Matches 0	Gaps 0
QY	541 AQTYSVAIFLGINLAAFIITVSYSGMFSVQSQTATATEIRNOYKKEMILAKREFFIVE				600
Db	1 aqysvaallglnlaafilivtsysmlysvngsatlateirnykkemilakrffivf				60
QY	601 TDAICWIPFIWVKFSLQVLEIPGRTTSWVVFIIPIINSALNPILYTLTRPFKEMIHRE				660
Db	61 tdaicwipfiwvkfslqlvlelpglitswvffilpinalnpllytltrpfkemihrf				120
QY	661 WYNRORKMSDKQKTTAPSTIWBEMPDLOEMPPELAKDULFTYPCENSLISQST				716
Db	121 wynyorkkmsdkxqgkltypasfciwemwplqempelimpdflfpcpcmsllsqst				176

RESULT	11
AAU18344	
ID	AAU18344 standard; Protein; 176 AA.
XX	
AC	AAU18344;
XX	
DT	21-NOV-2001 (first entry)
XX	
DE	Human endocrine polypeptide SEQ ID NO 299.
XX	
KW	Endocrine protein; human; mouse; rabbit; goat; horse; food additive;
KW	cat; dog; chicken; sheep; immunosuppressive; antiarthritic; vasotrophic;
KW	antirheumatic; antiproliferative; cyostatic; cardiant; neuroprotective;
KW	cerbroprotective; nootropic; antibacterial; virucide; fungicide; cancer;
KW	ophthalmological; vulnerary; gene therapy; autoimmune disease; neoplasm;
KW	hyperproliferative disorder; breast; liver; cardiovascular disorder;
KW	cerebrovascular disorder; nervous system disorder; bacterial infection;
KW	fungal infection; viral infection; ocular disorder; endocrine disorder;
KW	gastrointestinal disorder; renal disorder; respiratory disorder;
KW	wound healing; skin aging; organ transplantation; food preservative;
KW	tissue regeneration; anti-infertility.
XX	
OS	Homo sapiens.
XX	
PN	WO200155364-A2.
XX	
PD	02-AUG-2001.
XX	
PF	17-JAN-2001; 2001WO-US01308.
XX	
PR	31-JAN-2000; 2000US-0179065.
PR	04-FEB-2000; 2000US-0180628.
PR	24-FEB-2000; 2000US-0184664.
PR	02-MAR-2000; 2000US-0186350.
PR	16-MAR-2000; 2000US-0189874.
PR	17-MAR-2000; 2000US-0190076.
PR	18-APR-2000; 2000US-0198123.
PR	19-MAY-2000; 2000US-0205515.
PR	07-JUN-2000; 2000US-0209467.
PR	28-JUN-2000; 2000US-0214886.
PR	30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234323.
PR 21-SEP-2000; 2000US-0234374.
PR 25-SEP-2000; 2000US-0234597.
PR 25-SEP-2000; 2000US-0234598.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.

PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249246.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249329.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.
XX
XX
PI Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-451936/48.
DR N-PSDB; AAS29573.
XX
XX
PT Isolated polypeptide for treating, preventing and/ or prognosing
PT disorders of the endocrine system such as reproductive disorders,
PT endocrine cancers and also for testing and detection e.g. diagnosis -
XX
XX
PS Claim 11; SEQ ID No 299; 604pp; English.
XX
XX
Sequences AAU182-AU1807 represent endocrine polypeptides of the
CC invention. Endocrine polypeptides and their associated polynucleotides
CC are useful in the diagnosis, treatment and prevention of various types of
CC disorders in e.g. humans, mice, rabbits, goats, horses, cats, dogs,
CC chickens or sheep. A pathological condition can be determined by
CC determining the presence or absence of a mutation in an endocrine
CC polynucleotide. The treatable disorders include autoimmune diseases such

PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246509.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX
XX Rosen CA, Barash SC, Ruben SM;
PI WPI: 2001-488787/53.
XX N-PSDB; AAS30187.
DR
XX
XX New polynucleotides and polypeptides, useful for diagnosing, treating,
PT preventing or prognosing e.g. kidney, cardiovascular, blood,
PT electrolyte imbalance or neoplastic disorders, autoimmune diseases,
PT cancers
XX
XX Claim 1; SEQ ID No 105; 506bp; English.
XX
XX
XX The invention relates to novel nucleic acids and polypeptides useful for
CC diagnosing, treating, preventing and/or prognosing disorders related to
CC these polypeptides. The polynucleotides are especially useful in the
CC diagnosis, prognosis, prevention and/or treatment of diseases which
CC include kidney disorders (e.g. renal failure or nephritis),
CC cardiovascular disorders (e.g. hypertension or myocardial infarction),
CC blood disorders (e.g. anaemia or blood coagulation disorders),

CC electrolyte imbalance disorders (e.g. hyponatremia or hyperkalaemia),
CC neoplastic disorders (e.g. nephroma or renal cell cancer), autoimmune
CC diseases, cancers, inflammatory diseases, reproductive system
CC disorders, endocrine disorders, neural activity and neurological
CC disorders, wound healing and respiratory disorders. AAU18644-AAU18715
CC represent the novel human renal and cardiovascular-associated amino
CC acid sequences of the invention. Note: The sequence data for this patent
CC did not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at:
CC ftp.wipo.int/pub/published_pct_sequences.
XX
XX
SQ Sequence 176 AA;

Query Match 23.8%; Score 912; DB 22; Length 176;
Best Local Similarity 99.4%; Pred. No. 1.7e-80;
Matches 175; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 541 AQTSAIFLGINLAAEFIITVSYSGSMFYSHQSAITTEIRNQYKKEMILAKRFFIVF 600
Db 1 aqlysvaiflglnlaaeffiiivfsysmtfsvhsaaltateirngykkemilakrffivf 60
Qy 601 TDALCWIPFVYKFLSLQVEIPGRTTSMVYFIPIPSALNPILYTTTPPKEMHRRF 660
Db 61 tdaicwipifvvykflslqveipglttswwvifilpinalnpilyltltpfkemihrf 120
Qy 661 WYNVRQKRSKMSKQRTYAPSFYIWEWMPLOEMPPELKKPDLFTYPCMSLSIGST 716
Db 121 wynyrqksmdsxgqkyapsfiwemwplqempelmkpdlftypcemsllsigt 176

RESULT 13
AAU21646
ID AAU21646 standard; Protein; 176 AA.
XX
XX
AC AAU21646;
XX
XX 06-DEC-2001 (first entry)
DT
XX
XX Novel human neoplastic disease associated polypeptide #79.
DE
XX
XX Human: neoplastic disease associated polypeptide; cancer;
KW hyperproliferative disorder; neural disorder; immune system disorder;
KW muscular disorder; reproductive disorder; gastrointestinal disorder;
KW pulmonary disorder; cardiovascular disorder; renal disorder;
KW neuroprotective; cytostatic; anti inflammatory; vasotropic.
XX
OS Homo sapiens.
XX
XX W020015163-A1.
PN
XX
XX 02-AUG-2001.
PD
XX
XX 17-JAN-2001; 2001WO-US01358.
PF
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.

PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225216.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-022557.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229345.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 12-SEP-2000; 2000US-0232081.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234397.
PR 25-SEP-2000; 2000US-0234398.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0237945.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.

PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251899.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.
PA Rosen CA, Barash SC, Ruben SM;
PI WPI; 2001-465558/50.
XX N-PDB; AAS34845.
DR
XX
PT Novel polypeptides and polynucleotides useful as diagnostic reagents to
PT diagnose diseases or disorders associated with aberrant expression or
PT activity of polypeptides, and for treating cancers, rheumatoid
PT arthritis
XX
PS Claim 11; SEQ ID NO 373; 687pp; English.

XX The present invention relates to the isolation of novel human neoplastic
CC disease associated polypeptides, and cDNA (AAS34767-AAS35050) and DNA
CC sequences encoding for these polypeptides. The sequences of the
CC invention are useful in the diagnosis, treatment, prevention and/or
CC prognosis of disorders involving neoplastic disease such as
CC hyperproliferative disorders (e.g. leukemia, bone cancer, bladder
CC cancer, brain stem glioma, adult liver cancer, childhood cerebellar
CC astrocytoma, or Hodgkin's lymphoma). The sequences of the invention may
CC also be useful for treating other disorders such as neural disorders,
CC immune system disorders, muscular disorders, reproductive disorders,
CC gastrointestinal disorders, pulmonary disorders, cardiovascular disorders
CC and renal disorders. The polynucleotide sequences of the invention are
CC also useful in gene therapy. AAU21568-AAU21851 represent the novel human

CC neoplastic disease associated polypeptides of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 176 AA:

Query Match 23.8%; Score 912; DB 22; Length 176;
Best Local Similarity 99.4%; Pred. No. 1.7e-80;
Matches 175; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 541 AQLYSAIFGILNAAFIIVSGSMYSVHOSATATETIRNOVKEMIAKREFFIVF 600
Db 1 aqlysvaiflgilnaaflivsgsmysvhsatataetirnvkemiakrffivf 60
QY 601 TDALCWIPIFVNFELSLQVEIPGTTISWVIFILPINSALNPILVLTTRPEKEMIHRE 660
Db 61 tdaicwipifvfnfslqlveipgttitswvifilpnsalnpllvlttrpfkemihrf 120
QY 661 WYNYRQRKSDSGQKTYAPSFIVWEKMPIQEMPELMAKPDLTTPCEMSLISOST 716
Db 121 wynyrrqkmsdsgqkyapsfiwemwpiqempelmkpdltfypcemslisgst 176

RESULT 14

ID ABB04105 standard; Protein; 172 AA.

XX ABB04105;

DT 08-JAN-2002 (first entry)

DE Human musculoskeletal system related polypeptide SEQ ID NO 2052.

XX Cytostatic; immunosuppressive; nocotropic; neuroprotective; antiviral;
KM anti-allergic; hepatotropic; antidiabetic; anti-inflammatory; antilucer;
KM vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KM cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KM neurological disease; infection; human; secreted protein;
KM musculoskeletal system.

XX Homo sapiens.

PN WO200155367-A1.

PD 02-AUG-2001.

PF 17-JAN-2001; 2001WO-US01338.

XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.

PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 25-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.

PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
PI
XX
XX WPI; 2001-451937/48.
XX
XX N-PSDB; AAL35687.
XX
XX Isolated polypeptide for treating, preventing and/or prognosing
PT disorders related to the musculoskeletal system including
PT musculoskeletal cancers and also for testing and detection e.g.
PT diagnosis -
XX
XX
PS Claim 11: SEQ ID NO 2052; 781pp + Sequence Listing; English.

CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 172 AA;

Query Match 18.7%; Score 717.5; DB 22; Length 172;
Best Local Similarity 88.8%; Pred. No. 1,5e-61;
Matches 142; Conservative 3; Mismatches 14; Indels 1; Gaps 1;

QY 326 RMFRPLMNLISHIEKFKFOYGYAPVHRSCKPNTDGLISLENLASTIHFVFWVWSAVTC 385
|||||
Db 1 tmfrplmnlshlyfkfkfygygaphvrsckpntdglislenlasiqrfvfwvwsavtc 60
|||||
QY 386 FGNIFYICMRPYIRSENKLYAMSIIISLCADCMGYLFEVIGGFDLKFRGEYMKHAQLW- 444
|||||
Db 61 fgnifyicmrpyirsenklyamsiislcadcmgylyfvggfdlkfrgeykhqqlvd 120
|||||
QY 445 MESTHCOLVGSALILSTREVSULLITPLEKICYPPFR 484
| : ||||| : : ||||| : : |||||
Db 121 geyslxaxkdxlallstrevsxllvtlxltnkncldghpfr 160
|||||

RESULT 15
AAB41526
ID AAB41526 standard; Protein; 140 AA.
XX
XX AAB41526;
AC
XX
XX 08-FEB-2001 (first entry)
DT
XX
XX Human ORFX ORF1290 polypeptide sequence SEQ ID NO:2580.
DE
XX
XX
XX Human; open reading frame; ORFX; detection; cytosolic; hepatotropic;
KW vulnary; antiposrotic; antiparkinsonian; neurotropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antineoplastic;
KW antiviral; antibacterial; antifungal; antineoplastic; antihypertensive;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disease; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive.
XX
XX
XX Homo sapiens.
OS
XX
XX
XX WO200058473-A2.
PN
XX
XX 05-OCT-2000.
PD
XX
XX
XX 31-MAR-2000; 2000WO-US08621.
PF
XX
XX
XX 31-MAR-1999; 99US-0127607.
PR
XX
XX 02-APR-1999; 99US-0127636.
PR
XX
XX 05-APR-1999; 99US-0127728.
PR
XX
XX 30-MAR-2000; 2000US-0540763.
PR
XX
XX (CURA-) CURAGEN CORP.
PA
XX
XX
XX Shinkets RA, Leach M;
PI
XX
XX WPI; 2000-602362/57.
XX
XX N-PSDB; AAC75735.
XX
XX Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -
PT
XX
PS Claim 11: Page 1840-1841; 5507pp; English.

CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORF open reading frames 1 to 3161. The ORFs
CC sequences have activities such as: cytostatic; hepatotropic; vulnary;
CC antiproliferative; antiparkinsonian; nootropic; neuroprotective;
CC osteoplastic; anticonvulsant; antiarthritic; immunosuppressant;
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC antidiabetic; hypotensive; dermatological; immunosuppressive;
CC antiinflammatory; antibacterial; antiviral; antifungal; antineumatic;
CC antihypertensive; and antianemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an ORF-associated disorder. The
CC nucleic acids can be used to express ORF proteins in gene therapy
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.

XX Sequence 140 AA;

Query Match 17.0%; Score 650; DB 21; Length 140;

Best Local Similarity 99.2%; Pred. No. 4e-55; Mismatches 0; Gaps 0;

Matches 127; Conservative 0; Indels 1; Indels 0; Gaps 0;

OY 552 INIAFFIIVFSYSGMFSYHQSATATETIRNOVKKEMILAKRFFLVFTDALCWIPFV 611
DB 8 INIAFFIIVFSYSGMFSYHQSATATETIRNOVKKEMILAKRFFLVFTDALCWIPFV 67
OY 612 VKFLSLQVEIPGTISWVIFILPINSALNPILYTLTRPFKEMIHREWYNRQRKSM 671
DB 68 VKFLSLQVEIPGTISWVIFILPINSALNPILYTLTRPFKEMIHREWYNRQRKSM 127
OY 672 SKGQKTYA 679
DB 128 SKGQKTYA 135

Search completed: September 5, 2002, 08:25:16
Job time: 10676 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2002, 06:27:55 : Search time 66.25 Seconds
(without alignments)
266.193 Million cell updates/sec

Title: US-09-647-067-8
Perfect score: 3834
Sequence: 1 MTSGSVFFILIRKFKYSHG.....FTYPCMSLISOSTRLNSYS 722

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 2442594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCNUS_COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfilest.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	483.5	12.6	696	4	US-07-757-342D-4
2	481.5	12.6	692	4	US-07-757-342D-6
3	478	12.5	695	1	US-08-487-886-2
4	478	12.5	695	3	US-08-487-885-2
5	470.5	12.3	700	4	US-07-757-342D-3
6	469	12.2	674	4	US-07-757-342D-10
7	469	12.2	699	4	US-07-757-342D-2
8	449.5	11.7	611	4	US-07-757-342D-8
9	449.5	11.7	636	4	US-07-757-342D-7
10	441	11.5	795	4	US-07-741-453A-55
11	439	11.5	764	4	US-07-741-453A-59
12	438	11.4	764	4	US-07-741-453A-61
13	437	11.4	764	4	US-07-757-342D-5
14	430	11.2	764	4	US-07-741-453A-54
15	426.5	11.1	792	4	US-07-741-453A-56
16	395	10.3	764	4	US-07-741-453A-60
17	354.5	9.2	644	2	US-08-866-757-2
18	354.5	9.2	644	2	US-09-153-593-2
19	317.5	8.3	336	5	US-08-118-270-54
20	317.5	8.3	336	5	PCT-US93-08528-54
21	312	8.1	764	4	US-07-741-453A-29
22	312	7.9	764	4	US-08-190-802A-50
23	302	7.9	603	4	US-08-477-346-50
24	302	7.9	603	4	US-08-473-089-50
25	296	7.7	605	1	US-08-190-802A-49
26	296	7.7	605	4	US-08-477-346-49
27	296	7.7	605	4	US-08-473-089-49

28	294.5	7.7	332	1	US-08-118-270-53	Sequence 53, Appl
29	294.5	7.7	332	5	PCT-US93-08528-53	Sequence 53, Appl
30	293.5	7.7	605	4	US-09-063-950-5	Sequence 5, Appl
31	289	7.5	327	1	US-08-118-270-55	Sequence 55, Appl
32	289	7.5	327	5	PCT-US93-08528-55	Sequence 55, Appl
33	286.5	7.5	1480	3	US-09-191-647-7	Sequence 7, Appl
34	286.5	7.5	1480	4	US-09-540-245A-7	Sequence 7, Appl
35	286.5	7.5	1480	4	US-09-540-245A-7	Sequence 7, Appl
36	286.5	7.5	1480	5	PCT-US91-09055-2	Sequence 2, Appl
37	285.5	7.4	1480	4	US-09-182-024A-5	Sequence 5, Appl
38	285	7.4	1525	3	US-09-191-647-2	Sequence 2, Appl
39	285	7.4	1525	4	US-09-540-245A-2	Sequence 2, Appl
40	285	7.4	1525	4	US-09-540-153-2	Sequence 2, Appl
41	279	7.3	1523	4	US-09-182-024A-2	Sequence 2, Appl
42	278.5	7.3	353	3	US-08-986-485-6	Sequence 6, Appl
43	273	7.1	1091	3	US-08-986-485-5	Sequence 5, Appl
44	267.5	7.0	1101	3	US-08-986-485-2	Sequence 2, Appl
45	259.5	6.8	649	4	US-09-188-930-305	Sequence 305, App

ALIGNMENTS

RESULT 1
US-07-757-342D-4
Sequence 4, Application US/0757342D
Patent No. 6218509
GENERAL INFORMATION:
APPLICANT: IGARASHI, Masao
MINESIGHT, Takashi
NAKAMURA, Kazuo
TITLE OF INVENTION: PROTEIN, DNA AND USE THEREOF
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESS: DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS &
CUSHMAN
STREET: 130 Water Street
CITY: Boston
STATE: Massachusetts
COUNTRY: US
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,342D
FILING DATE: 10-SEP-1991
CLASSIFICATION: <unknown>
ATTORNEY/AGENT INFORMATION:
NAME: BUCKLEY, Linda M.
REGISTRATION NUMBER: 31003
REFERENCE/DOCKET NUMBER: 41226
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)523-3400
TELEFAX: (617)523-6440
TELEX: 200291 STRE UR
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 696 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-07-757-342D-4
Query Match 12.6%; Score 483.5; DB 4; Length 696;
Best Local Similarity 23.3%; Pred. No. 1.2e+30;
Matches 144; Conservative 133; Mismatches 261; Gaps 17;
QY 87 PSVSSVTVAMSLQWNLIRKLPPDCKRNVHDLQKLDL-QNNKITSISIVAFGLNSLTGLXY 145

```

Db      45  PPRAGLSRLSTLYTLTKIVIPSOAFRGNGEVEVKKIEQSDSELEKIEANAPDNLLTSEL 104
Qy      146  LSHNR-ITFLKPGVEDLHRLFEMLIENHLSRSPPL- FYGLNSLLIYAMNVLTSL 202
Db      105  IONKNLVIIEGATNPLRLKYSLICWTGIRKLPDYTKIFSSSEFNLTLECDLHITTV 164
Qy      203  PDKPLCQHPRLHMLDLEGNHILNRLNTFISCSNLYVLVARKKKIHNLNNTFPAPLOKL 262
Db      165  PANAFQGNMNSITLKLKVGNEFEELQSHAFNGTLLISELKENHAKMKHNDARGARCP 224
Qy      263  DELDGSKKIEMLPPLIKDKLEL---SQLNLSNPLOKIOANOPDYLVKIKLSLEGIE 319
Db      225  SILDISSYKLOALPESYGLESYOTLIATSSYSKLPLPSREKFTNLLDATALYPSHCC- 280
Qy      320  ISNIOGRFRLP---MNLSHYFRKFO----- 343
Db      281  -----AFRLNLPKQGNFESIFIRKNSKQCESTARPNNETLKSATFASELSDMWDY 333
Qy      344  -YCGAPHVNSCKPNTDGISSLENLASIIQRFVWVWVSAVTCGNIFVICMPYRSEN 402
Db      334  GFC--SPKTLQCADEPDFAFNCEIDMGDFRLVLIINILIAINGVTVLEV--LTSHY 389
Qy      403  KLYANSII--SLCADCCLMGTYLEVIGSFDLKRGEYKHNKQIMMESTHQVLSLAIS 466
Db      390  KLTVPRLMCNLSRFDPCMGTLILILIASVDQOTGYIYHAIIDQTNCGSVAGFTVFA 449
Qy      461  TEVSALLTFLLEKYICIVYFPCVRPKCR-TITVILIMTIGTVAFIPLSNKEFFK 519
Db      450  SELSYVTLTVLEMRHTITVAIDQDKLRLRHAIPIMLGWLSTLIAMPVG----V 505
Qy      520  NYGNGVCPFLHSEDETSIAQIYSAIFLGINLAFIITVEYSGMEFYSVHOSATTAT 579
Db      506  SSKMVSYSICLDP--DVETTLISQVILITLIL-LNVVAFIILICACYIKIYFVONPELMAT 561
Qy      580  EIRNOVKEMILAKREFEIVTDLCAWPIPVVFLSLQVEIPGITSWVIFIL--PI 637
Db      562  -----NDQTIARKMAVLITFDTCAPISFPAISAAKAPPL-TVNSKVLVLYEFPV 614
Qy      638  NSALNPILYITLTTRPK 654
Db      615  NSCANPFLYATFTKFR 631

RESULT 2
US-07-757-342D-6
Sequence 6, Application US/07757342D
Patent No. 6218509
GENERAL INFORMATION:
APPLICANT: IGARASHI, Masao
MINEGISHI, Takashi
NAKAMURA, Kazuo
TITLE OF INVENTION: PROTEIN, DNA AND USE THEREOF
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS & CUSHMAN
STREET: 130 Water Street
CITY: Boston
STATE: Massachusetts
COUNTRY: US
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,342D
FILING DATE: 10-Sep-1991
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: BUCKLEY, Linda M

```

```

?      REGISTRATION NUMBER: 31003
?      REFERENCE/DOCKET NUMBER: 41226
?      TELECOMMUNICATION INFORMATION:
?      TELEPHONE: (617)523-3400
?      TELEFAX: (617)323-6440
?      TELEX: 200291 STRE UR
?
?      INFORMATION FOR SEQ ID NO: 6:
?      SEQUENCE CHARACTERISTICS:
?          LENGTH: 692 amino acids
?          TYPE: amino acid
?          TOPOLOGY: linear
?      MOLECULE TYPE: protein
?      SEQUENCE DESCRIPTION: SEQ ID NO: 6
US-07-757-342D-6

```

Query Match	12.6%;	Score 481.5;	DB 4;	Length 692;
Best Local Similarity	22.7%;	Pred. No. 1.8e-30;		
Matches 169;	Conservative 140;	Mismatches 272;	Indels 165;	Gaps 28;

QY	35	GNITTCPLHCHGVNDGQADBNCCVVVLCQOMSLPGLELDMKPFVSVP-SVSN	93
Db	13	GTGSCCHMLCHCSN-----RVLQDSKV-----ZELPDLPRNA	48
QY	94	TAMSLQMWLIRKLPPDCKNTHDLOKLDLONNKTTSTIYAFRLNSITKLYLSHRITF	155
Db	49	IELRVLTKLRVIRPGSFAGGDEKLEISONDVLE-----	85
QY	154	LKPGVEEDLHREMLIID-NHLRISPPRYGNSILVLMNNVLRPLDRLCOHMP	212
Db	86	IEADVSNLPLHREIRKANNLLYINPEAFONLPSLRYLILSTIGIKHLPAYHKIOSLO	145
QY	213	RLHMLDEGN-HIHNLRNLFEFICS-NLTVLMRKNKINHNLNENTFAPLOKDELUDG-S	269
Db	146	KV-LDDIODINIHIVARNSPMGLSPESVILMLSKNGIEIHNCAFNGOTQ-LDELNLSDN	203
QY	270	NKIEMLPLIRKDKLSQLMLSTNPLOKIOANOFDYLVLKLSLSLGIE-----	319
Db	204	NNLEELPNDVFOGASGPVILDISRTKXHSIPNHGLEMLKLRARSTYRLKLPMLDKVPT	263
QY	320	-----ISNQOQM-FRPLMNL-----	334
Db	264	LMESLVIPSHCCAPANKRKROISELHPICNNSILRQDIDMTQIGDORVSLIDDEPSYK	323
QY	335	-SHYFKKKFOY--GQYAPHVRSCKPNTDGISSLENLASHIORYFVWVSATVCFGNFV	391
Db	324	GSDMYNNEFDLDCNEVYD-TCPKRPDAPNCPGDINGYLNRLVILFISLATGNTYV	362
QY	392	TCMRPYIRSENKLYAMSLI-SLCCADCLMGILYLVJGFDLKRGEYKRNHQAOLMESTH	449
Db	383	LVV--LTSQYKLLVPRFLMCLNAFADLCIGIYLLLASVDIHKSGYHNVAIDWQGAG	440
QY	450	COLVGSIAISTEYSVLLTFLTEKICYICIVPPRCVAPGCR-----TITVLILMTG	504
Db	441	CDAGGFVTFVASELSVYTLTATLERMHTTHAOL-----ECKQVLRAHASVWVLGWTFA	466
QY	505	FIVAFIPLTSNKEFKKNYGTGVOEPLHSDEPSESIGAQIVSVAFLSINLAFFIIVFSY	564
Db	497	FAALFPL-----FGISSTYMKYSICLPM--DISPLSQLTYMAL-LVLNLAFAVFIGCY	548
QY	565	GSMEYSVHOSATITATELINQVKKEMILAKRFFVIFVLDALCWLIPFVFKFLSLQVEIPG	624
Db	549	THYLVANPNPIVS-----SSDTKAKRAKATLIFDFDPLCMAPISFALSASLKVLI-	601
QY	625	TITSWVVFIL--PINSALNFIPLYTLTPRPEK--MIHRF-----W	661
Db	602	TYSKAKILLVLYPINSKANPFLYAIFTKNBRDFILLSKFGCYEQMAQIYRTETSSAT	661
QY	662	YNYRQKSMDSKCGOKTAPSPITWEM	687
Db	662	HNFAKRSCHSAPRV-TNSVYVPL	686

```

RESULT 3
US-08-487-886-2
: Sequence 2, Application US/08487886
: Patent No. 5744448
: GENERAL INFORMATION:
: APPLICANT: Kelton, Christie Ann
: APPLICANT: Schweickhardt, Rene Lynn
: APPLICANT: Cheng, Shirley Vui Yen
: APPLICANT: Nugent, No. 5744448een Patrice
: TITLE OF INVENTION: Human Follicle Stimulating
: TITLE OF INVENTION: Hormone Receptor
: NUMBER OF SEQUENCES: 2
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Stephan P. Williams,
: ADDRESSEE: Ares-Serono, Inc.
: STREET: Exchange Place, 37th floor
: CITY: Boston
: STATE: MA
: COUNTRY: USA
: ZIP: 02109
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 3.5" diskette, 1.44 MB, high density
: COMPUTER: IBM PS/2, model 55 SX
: OPERATING SYSTEM: MS-DOS version 4.0
: SOFTWARE: VAX/VMS Mass11 via Kermit to IBM MS-DOS
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/487,886
: FILING DATE: 07-JUN-1995
: CLASSIFICATION: 514
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 07/670,085
: FILING DATE: 15-MAR-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: Williams, Stephan P.
: REGISTRATION NUMBER: 28546
: REFERENCE/DOCKET NUMBER: US/252
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (617) 723-1300
: TELEFAX: (617) 723-8923
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 695
: TYPE: Amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: FEATURE:
: NAME/KEY: signal sequence
: LOCATION: -17 to -1
: IDENTIFICATION METHOD: hydrophobic
: FEATURE:
: NAME/KEY: putative amino-terminal extracellular domain
: LOCATION: 1 to 349
: IDENTIFICATION METHOD: similarity with other
: IDENTIFICATION METHOD: dimeric glycoprotein receptor extracellular
: IDENTIFICATION METHOD: domains, hydrophilic
: FEATURE:
: NAME/KEY: transmembrane domain
: LOCATION: 350 to 613
: IDENTIFICATION METHOD: similarity to other G
: IDENTIFICATION METHOD: protein-coupled receptor transmembrane domains
: FEATURE:
: NAME/KEY: putative transmembrane region I
: LOCATION: 350 to 370
: IDENTIFICATION METHOD: similarity to other G
: IDENTIFICATION METHOD: protein-coupled receptor transmembrane regions,
: IDENTIFICATION METHOD: about 20-23 amino acids in length
: FEATURE:
: NAME/KEY: putative transmembrane region II
: LOCATION: 382 to 404
: IDENTIFICATION METHOD: similarity to other G
: IDENTIFICATION METHOD: protein-coupled receptor transmembrane regions,
: IDENTIFICATION METHOD: hydrophobic, about 20-23 amino acids in length
: FEATURE:

```

```

: NAME/KEY: putative transmembrane region III
: LOCATION: 427 to 448
: IDENTIFICATION METHOD: similarity to other G
: IDENTIFICATION METHOD: protein-coupled receptor transmembrane regions,
: IDENTIFICATION METHOD: about 20-23 amino acids in length
: FEATURE:
: NAME/KEY: putative transmembrane region IV
: LOCATION: 469 to 491
: IDENTIFICATION METHOD: similarity to other G
: IDENTIFICATION METHOD: protein-coupled receptor transmembrane regions,
: IDENTIFICATION METHOD: about 20-23 amino acids in length
: FEATURE:
: NAME/KEY: putative transmembrane region V
: LOCATION: 512 to 533
: IDENTIFICATION METHOD: similarity to other G
: IDENTIFICATION METHOD: protein-coupled receptor transmembrane regions,
: IDENTIFICATION METHOD: about 20-23 amino acids in length
: FEATURE:
: NAME/KEY: putative transmembrane region VI
: LOCATION: 557 to 580
: IDENTIFICATION METHOD: similarity to other G
: IDENTIFICATION METHOD: protein-coupled receptor transmembrane regions,
: IDENTIFICATION METHOD: about 20-23 amino acids in length
: FEATURE:
: NAME/KEY: putative transmembrane region VII
: LOCATION: 592 to 613
: IDENTIFICATION METHOD: similarity to other G
: IDENTIFICATION METHOD: protein-coupled receptor transmembrane regions,
: IDENTIFICATION METHOD: about 20-23 amino acids in length
: FEATURE:
: NAME/KEY: putative carboxy-terminal intracellular
: LOCATION: 614 to 678
: US-08-487-886-2

```

```

Query Match 12.5%; Score 478; DB 1; Length 695;
Best Local Similarity 23.1%; Pred. No. 3.4e-30;
Matches 160; Conservative 129; Mismatches 250; Indels 154; Gaps 26;

QY 40 CLPOLHNGVDCGNOADENCYVVLCOGMSLPGLEIDWKKPPTSVPS-VSSVWTAMSL 98
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 18 CHHRICHGCSN-----RVFLCQ-----ESKYTEIPSDLPRAAIELRF 53

QY 99 QWNILIRKLPDPCFKNYHDLOKIDLONNKITSISYAFRGSLTKLYSHNRITFLKPCV 158
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 54 VLRKLRVIOKGFSGFGDLKEIKETISONDVLEV-----IEADV 90

QY 159 FEDLHRLFWLITFD-NHLSTRISPTFGINSLLILVLMNNVLRPLPKQHPRLH-- 215
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 91 FSNLPKLEHIEIKANNILYINPAFONLPNLOYLLISNIGIKHLPD-----VKIHSL 144

QY 216 --WIDLEGN-HIHLNRLTFISCS-NLTVLMRKNNIHLNENTFAPLOKIDELDLS-S 269
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 145 QKVLDDIQDNNINHTIERNSFVGLSPFESVILMLKNGIOEIHNCFAFGTQ-LDELINSDN 203

QY 270 NKIENLPLIRKDKELSQLNLSTNPLOKIOANOPDLYLAKLSLSGIE----- 319
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 204 NNLEELPDVPHGASGPVYLDISRTIHSPSYGIENLKLRASTYLNKRLPTLEKLYA 263

QY 320 -----ISNIOGRM--FRPLMNL----- 335
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 264 LMEASLYPSSHCAFPANMRROISLHPICNKSILRQEDVYMTQRRGSSLAEDNESSYS 323

QY 336 --HIYEKKFOY--CGYAPHVRSCKPNTDGISSLENTLAIITQVFWVSAVTCFGNIF 390
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 324 RGFMTYTEFDYDLCNEVDV-TCSPKPDAPNCPEDIMGNYILRLWFIITLITGNI 382

QY 391 VTCMRPYIRSNKLYASII--SLCADCLMGITLFLVYIGGDDLFRGEXYKNAHQLMEST 448
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 383 VLVI--LTTSQYKLTVPFELMCNLAFADLGIGIYLLILASVDHTKSKOYHNAYADWGTGA 440

QY 449 HQQLVGSIALISTEVSVLLTFLLEKYICIVPFRGVPRGCKR-----TTVLILMIT 503

```

[illegible][illegible]

Db 204 NNLELPNDVHGASGPVILIDISTRILHSLPSYGLNKKIKARSTYVLLKPLLEKLYA 263
QY 320 -----ISNIQRM--FRPLMNL----- 335
Db 264 LMEASLYPSHCAFAWMRRIOISELHPICNKSILROEVDYMTQRTGORSIAEDNESSYS 323
QY 336 ---HLYKKFPQY--CGVAPHRSCKPNTDGISSLLENLALSIQVFWVWVAVVCFQWIF 390
Db 324 RQDMTTEFEYDLCNEVVD--TCSPPDAPNCPEDIMGNILVLMFISILAITGNI 382
QY 391 VICMRPIRSENKLYAMSII--SLCADCLMGITLFLVIGFDLKFGEYKHAQLMEST 448
Db 383 VLVI--LTTSGYKLTVPFLMCLNLAFLDLGIGYILLILASDIHTKSGYHNYALDMQGA 440
QY 449 HQVLGSLAISTEVSALLFTFLLEKTCIYVPRVCYRQCR---TTVLLMIT 503
Db 441 GCDAGFEYVASELSYVTLTAITLERMHTTHAMQL---DCKVQLRHASVVMGWIF 496
QY 504 GEIYAFPLSKKEPKNYGNGVCPPLSHSEDSIGAOIYVAIFGINLAFFIIVFS 563
Db 497 AFPAALPFI---FGISYMKVSTCLPM--DIDPLSQLYMSL--LVNLVLAFLVIGCC 548
QY 564 YGSMFYVHOSAITATEIRNOVKKEMILAKRFFIVFTDALCWIPIFYVKFLSLQVEIP 623
Db 549 YIHLYLVFNBNIVSS-----SSDTRIARKMAMLIFTDFICMAPISFEAISASIKVPLI 602
QY 624 GTITSWVIFIL--PINSALNPILYTLTTRPK 654
Db 603 -TVSKAILLVLPINSCANPFLYAITKFR 634

RESULT 5

US-07-757-342D-3

; Sequence 3, Application US/07757342D
; Patent No. 6218509

; GENERAL INFORMATION:

; APPLICANT: IGARASHI, Masao

; MINEGISHI, Takashi

; NAKAMURA, Kazuo

; TITLE OF INVENTION: PROTEIN, DNA AND USE THEREOF

; NUMBER OF SEQUENCES: 10

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS &

; CUSHMAN

; STREET: 130 Water Street

; CITY: Boston

; STATE: Massachusetts

; COUNTRY: US

; ZIP: 02109

; COMPUTER READABLE FORM:

; MEDIUM TYPE: floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/07/757,342D

; FILING DATE: 10-Sep-1991

; CLASSIFICATION: <unknown>

; ATTORNEY/AGENT INFORMATION:

; NAME: BUCKLEY, Linda M.

; REGISTRATION NUMBER: 31003

; REFERENCE/DOCKET NUMBER: 41226

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (617)523-3400

; TELEFAX: (617)523-6440

; TELEX: 200291 STRE UR

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 700 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-07-757-342D-3

Query Match 12.3%; Score 470.5; DB 4; Length 700;

Best Local Similarity 22.8%; Pred. No. 1.3e-29;

Matches 149; Conservative 135; Mismatches 233; Indels 137; Gaps 23;

QY 87 PVSNNVTAMSLQNNLRKLPPDCKNHDQKIDL-QNNKITSISIAFGLNSLTKLY 145
Db 49 PGPRAGLARLSLTLYPVVIPSQAFRGVLEVKIEISQSDSLERENAFVNLNLSLL 108
QY 146 LSHNR-LTFELKPGVEDHRLLEMLIEDN-----HLSRISPTPEYGLNSLILYAMNV 198
Db 109 IONTKNLIYIEPGAFNLPRIKYSICTGTIRLPDYTKISSFE-----NFILECDNH 164
QY 199 LTRLPDKPLCOHMPRLHMLDLEGNHINLRNLJFISCSNLTVLVMRKKNHNLNENFPAP 258
Db 165 ITTIPGNAFQGMNNSVTLKLYNGOFEEVQSHAF-----NGTTLISLE-----LKENIY-- 213
QY 259 LQKIDE-----LDLSKNKIEENPLPIFDKLSQLNLSYNPIQIQANQPFYL 307
Db 214 LEKMSGAFOGATGPSLIDISTTKLQALPS-----HGLSIIQT-----L 252
QY 308 VKLSLSLEGIEISNIQRMFRPLM-----NLSHYFKKKFQ--- 343
Db 253 IALSYSLSKTLR---SKKFTSLVATLTYPSHOCARNLPKKQONFSFIFENFSKQC 308
QY 344 -----YCGVAPHRSCKPNTDGISSLLENLALSIQ 373
Db 309 ESTVRKADNETLYSAIFEEENELSGWDYDYGFC--SPKYLQCAPEDAPNCPEDIMGNVAF 366
QY 374 RFWVWVSAVTCFGNIFVFCMRPIRSENKLYAMSIIISLCADCLMGITLFLVIGFDLKF 433
Db 367 RVLWLNILIIILFGLVLYLVLSRYKLYVPRFLMCLSTRADFCMGYLLILASVSDQT 426
QY 434 RGEYKHAQLMESTHCOLVLSALISTEVSALLFTFLLEKTCIYVPRVCYRQCR- 492
Db 427 KGOYNNHAIIDMOTSGCGAAGFTYVASELSYVTLTTLERMHTIYAVQDQRLRH 486
QY 493 TTVVILIMITGFIYAFPLSKKEPKNYGNGVCPPLSHSEDSIGAOIYVAIFGI 552
Db 487 AIPIMGLMFLSTLATPMLVG---ISNYMKVS-ICLPM--DVESTLSOVIISIL-L 538
QY 553 NLAFFIIVSGSMFYVHOSAITATEIRNOVKKEMILAKRFFIVFTDALCWIPIFYV 612
Db 539 NVAVFVVICACIRITFAVQNPDELAP-----NDOTIAKMAIILFTDFICMAPISFF 592
QY 613 KFLSLQVEIPQTTTSWVIFIL--PINSALNPILYTLTTRPKFKE---MIHRF 660
Db 593 AISAFAKVPFI-TVNNSKILLVLPVNSCANPFLYAITFAFQDFDILLLSRF 645

RESULT 6
US-07-757-342D-10
; Sequence 10, Application US/07757342D
; Patent No. 6218509
; GENERAL INFORMATION:
; APPLICANT: IGARASHI, Masao
; MINEGISHI, Takashi
; NAKAMURA, Kazuo
; TITLE OF INVENTION: PROTEIN, DNA AND USE THEREOF
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS &
; CUSHMAN
; STREET: 130 Water Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: US
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk

```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,342D
; FILING DATE: 10-Sep-1991
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: BUCKLEY, Linda M.
; REGISTRATION NUMBER: 31003
; REFERENCE/DOCKET NUMBER: 41226
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)523-3400
; TELEFAX: (617)523-6440
; TELEX: 200291 STRE UR
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 674 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 10:
US-07-757-342D-10

```

```

Query Match          12.2%; Score 469; DB 4; Length 674;
Best Local Similarity 23.0%; Pred. No. 1.7e-29;
Matches 144; Conservative 140; Mismatches 277; Indels 64; Gaps 18;

```

```

QY 67 CQCMSLPGLDMMKPTSPVSSVNTAMSLOWNLRLKPPDCKFYHDLQKLDL-QNN 125
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 9 CNCVPDGLR-----CPGPTAGLTRSLAVLPVKYIPSAQFRLNEVIKIEISQID 59
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 126 KITSISYAFRGINSILTKILSHNR-ITFLKPGVFEDLHLEWLIIDNHLRSISPT-- 182
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 60 SLERIEANADNLINSEILQNTKRLRYIEPGAFLNLPGLKYLISICNTGIRKFPDVTKY 119
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 183 FYGLNSILILVLMANNVTRLPDKPLCOHMRPLMLDLEGNHINRLRLTIFSCNLTVLV 242
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 120 FSSFSNFIELICDNLHTTIPGNAFOGMNNEVTLKLYGNGFEVOSHAF-NGTTLTSL 178
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 243 MRKN-KINHLNENTFAPLOKLDLDSNKTENLPLIFDOLKEL---SOLNLSYNP10K 298
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 179 LKEVHLLEKHNNGAFRGATGPKTIDISTTKLOALPSYGLSIOQLATSSYLKLPSSRE 238
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 299 IOAN-----QFDYLVK---LKSLSLEGIEISNIQOMRFRPLM 332
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 239 TFVNLLEATLTPSHCCAFRNLPKRONFHSISENFSKOCESIVRKVSN--KTLYSML 296
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 333 NLSHIYFKKFOYCGYAPHVRSCKPNTDGLSLENLILASIIQRFVWVAVTCFGNIFVI 392
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 297 AESLSGMDVEYGFCLPKTRCAPEPDAPNCPEDIMGYDLRLVILWILNITLAINGNMTVL 356
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 393 CMRPYIHNENKLYAMSLISCCADCLMGILYLVYIGFDLKFGRGYNKHQOLMESTHCQL 452
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 357 FVLITSTYKLTVPFRLMCNLSFADFQMGVLLILLASVDSQTKGQYHNAIDWQSGSCST 416
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 453 VGSLSLITSEVSVLLFELFLKLYICIVYPERCVRPGKCF-TITVLLIWLITGIVAFIP 511
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 417 AGFFTYFASLSTYTLVITLERNHTITTAIHLDQKLRHALLIMGWLFPSSLIMLP 476
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 512 LSNKEFFKNTYNGVCFPLHSEDTESIGAQISVAIFLGINLAELTIIVFSYSGSMFYSV 571
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 477 LVG---VSNNMKYS--IcfPM---DVEITLSQVYILITLI-LNVAFRITLACAYIKIIFAV 528
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 572 HQSIAITATELRNOYKKMILAKRPFIVPTDALCWPIDFVVKLSLQVEIPGTTISWV 631
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 529 RNPELMAT-----NKDTKIAKKMAILIFDTCMAPISFEFAISAARKVDLI-TVTSKY 581
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 632 IFIL--PINSALNPILYTLTTRPRK 654
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 582 LVLVFPINSCANPELAIPTKTFQ 606
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

RESULT 7
US-07-757-342D-2
; Sequence 2, Application US/0757342D
; Patent No. 6218509
; GENERAL INFORMATION:
; APPLICANT: IGARASHI, Masao
; NAKAMURA, Kazuo
; TITLE OF INVENTION: PROTEIN, DNA AND USE THEREOF
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DAVID G. CONLIN, DIKE, BRONSTEIN, ROBERTS &
; CUSHMAN
; STREET: 130 Water Street
; City: Boston
; STATE: Massachusetts
; COUNTRY: US
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,342D
; FILING DATE: 10-Sep-1991
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: BUCKLEY, Linda M.
; REGISTRATION NUMBER: 31003
; REFERENCE/DOCKET NUMBER: 41226
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)523-3400
; TELEFAX: (617)523-6440
; TELEX: 200291 STRE UR
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 699 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-07-757-342D-2

Query Match          12.2%; Score 469; DB 4; Length 699;
Best Local Similarity 23.0%; Pred. No. 1.8e-29;
Matches 144; Conservative 140; Mismatches 277; Indels 64; Gaps 18;

```


TELEX: 200291 STRE UR
 INFORMATION FOR SEQ ID NO: 7:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 636 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 SEQUENCE DESCRIPTION: SEQ ID NO: 7:
 US-07-757-342D-7

Query Match 11.7%; Score 449.5; DB 4; Length 636;
 Best Local Similarity 22.5%; Pred. No. 5.7e-28;
 Matches 134; Conservative 134; Mismatches 259; Indels 69; Gaps 16;

QY 67 CQCSLPLGLDMMKPTSPVSSNTYAMSLQNLIRKLPDQFKNYHDLQKLDLQNN 125
 DB 34 CMCVPGDALR-----CGPPTAGLRSLALPVKAVIPSOAFRGLINEYKIEISQID 84
 QY 126 KITSISYAFRGSLNLRKLYLSNR-ITFLKPGVEDLHRLLEMLIEDNHLRSISPT-- 182
 DB 85 SLERIEANAFDNLNLSEILIQNTKRLYIEGAFINLPGLKYSICNTGIRKRPDVTKV 144
 QY 183 FYGINSLLVLMANNVTRLPDKPLCOHMRPLHWDLEGNHNLRLFTISCSNLTFLV 242
 DB 145 FSSRSNFIIECDMLHTTTPGNAFOGMNDSYTLKLYGNFEEVQSHAF-NGTTLSLE 203
 QY 243 MRKN-KINHNLNENFAPLOKDELDELGSNKIENLPPLIFDKELSQLNLSYNIQKIQ 301
 DB 204 LKEVNHLEKHNKNGAFRGATGKTONFSHSISENFSKOCESIVRKVSKNTLSSMLASEL 263
 QY 302 NQEDYLVKLSLSLEGIEISNIQRMFRPLMNLSHYFKFQYCGYAPHVASCCKPNTDGI 361
 DB 264 SGWDY-----EXGFC--LPKPRCAPPEDAF 287
 QY 362 SLENLASTIIOQFVWVSAVTCFQNIIFYICAMPYIRSEKRLYAMSIISLCCADLMGI 421
 DB 288 NPCEDINGYDFLRVLIWLIIMGMVLEVLITSRKYLTPRFIMCINLSPDFCMGL 347
 QY 422 YLIFGFDLFRGEYKHAOLMESTHCOLVGSLLISTEVSALLTFPLEKICYIC 481
 DB 348 YLLILASVDQTKQYNNHADMOTGSCGAGFTYFASLSLYTTLVITLLEMHITTY 407
 QY 482 PRCVPRGKCR-TITVLLITWITGIVAFPLSNKEFFKNYGTNGVCFPLHSDTESIG 540
 DB 408 AIHLDQRLRLHAILMLGWLFSLLIAMLPLVG--VSNNMKVS-ICFPV--DVEETL 460
 QY 541 AOIYSVAFIIGINLAARITIVESGSMFYHQSATITATERNVKKEMILAKREFFIV 600
 DB 461 SOYITLITL-LNVVAFITICACIKTYFARNDELMAI-----NDOTKIAKKMALILIF 513
 QY 601 TDALCWPIFEVVFSLAQVEIPETITSWVIFIL--PINSALNPILYTLTRPCK 654
 DB 514 TDFCMAPISEPAISAFKVPPLI-TVTNSKVLVLVFPINSCANPFLXALFTKRFQ 568

RESULT 10
 US-07-741-453A-55
 Sequence 55, Application us/07741453A
 Patent No. 6228597
 GENERAL INFORMATION:
 APPLICANT: PARMENTIER, MARC
 APPLICANT: LIBERT, FREDERIC
 APPLICANT: DUMONT, JACQUES
 APPLICANT: VASSART, GILBERT
 TITLE OF INVENTION: POLYPEPTIDES HAVING THYROTROPIN-RECEPTOR
 TITLE OF INVENTION: ACTIVITY, NUCLEIC ACID SEQUENCES CODING FOR SUCH RECEPTORS
 NUMBER OF SEQUENCES: 62
 CORRESPONDENCE ADDRESSES:
 ADDRESSEE: CUSHMAN DARBAY & CUSHMAN
 STREET: 1615 L STREET, N.W.
 CITY: WASHINGTON, D.C.

COUNTRY: U.S.A.
 ZIP: 20036
 COMPUTER READABLE FORM:
 MEDIUM TYPE: disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/07/741,453A
 FILING DATE: 19911015
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: KOKULIS, PAUL N.
 REGISTRATION NUMBER: 16773
 REFERENCE/DOCKET NUMBER: 91913/1107/US/ST
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202) 861-3000
 TELEFAX: (202) 822-0944
 TELEX: 6714627 CUSH
 INFORMATION FOR SEQ ID NO: 55:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 795 amino acids
 TYPE: AMINO ACID
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-07-741-453A-55

Query Match 11.5%; Score 441; DB 4; Length 795;
 Best Local Similarity 22.2%; Pred. No. 3.6e-27;
 Matches 157; Conservative 131; Mismatches 230; Indels 190; Gaps 29;

QY 87 PVSANTYNTAMSLQNLIRKLPDQFKNYHDLQKLDLQNNKIT-----SISYAFRGLN 139
 DB 61 PGPAGIARSLTYLPVKVIPSQAFRGLNEVYKIEISQSLERATHCGRIEANAFLN 120
 QY 140 SLKRLYLSNR-ITFLKPGVEDLHRLLEMLIEDN-----HLSRISPPRYGL 186
 DB 121 NLSELLIQNTKRLYIEGAFINLPRLKYSICNTGIRATHCGLPDVTKISSSEF--- 177
 QY 187 NSLILVLMANNVTRLPDKPLCOHMRPLHWDLEGN-----HHNLRNLFTISCSNLT 240
 DB 178 -NFLIECDMLHTTTPGNAFOGMNDSYTLKLYGNFEEVQSHAF--NGTT 232
 QY 241 LVMRKNKINHNLNENFAPLOKDE-----IDLGSNKIENLPPLIFDKELSQL 289
 DB 233 LISIE-----LKENYI--LEKMSGARQATGSILDISITKLD-----ARATHC 275
 QY 290 NLSYNIQKIQANQFDYLVKLSLSLEGIEISNIQRMFRPLM----- 332
 DB 276 GLPSHGLSISQIOT-----LIALSSYSIKTLP-----SKKFTSLVAVLTUPSHCCAFRNL 326
 QY 333 -----NLSHYTKRFQ----- 343
 DB 327 RATHCGKKEQNFSESIENFSKOCESIVRRADNETIYSAIFENELSGMDRATHCGDYG 386
 QY 344 ---YCGYAPHVRSCKPPTDGISSLENLSTIIOREVWVSAVTCFQNIIFYICMRP----- 396
 DB 387 RATHCGSPFTIQCAPPDPAFNCEDIMGAFRLVLIWLIINILAIENLTLVLYVRATHCG 446
 QY 397 YRSENKLYAMSI--SLCCADLMGIYLFVIGFDLFRGEYKHAOLMESTHCOLVG 454
 DB 447 LILSRKYLTVPRFLMCLSPADFCMGYLLILASVDSQTKQYNNHADM-RATHCGQGTG 505
 QY 455 S-----LAILTEVSALLTFPLEKICYIYP-----PRCVPRGKRTIYVLL 499
 DB 506 SGCGAGFTFVASELSVYTLTLERMHTIYAVQLDQKRLRNRATHC-GIPMLG 564
 QY 500 IWTGIVAFIPLSNKEFFKNYGTNGVCFPLHSDTESIGAOIYSVAIFL-----GINL 554
 DB 565 GMLPSTLIATMPLVG--ISNNMKVS-ICLPM--DVESTLSOYVILSLIRATHCGGLNV 617

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2002, 08:16:10 ; Search time 61.54 Seconds
(without alignments)
1127.340 Million cell updates/sec

Title: US-09-647-067-8

Perfect score: 3834
Sequence: 1 MHSQSVFFILLFKYFSGH.....FTYPCMSLISQSTRLNST 722

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	832	21.7	113.5	S40241	G protein-coupled
2	501	13.1	695	JC1493	folliotropin recept
3	498	13.0	695	I45896	follicle stimulat
4	490	12.8	907	JG0193	G protein-coupled
5	484.5	12.6	694	JC2237	folliotropin recept
6	482.5	12.6	694	JC4301	folliotropin recept
7	481.5	12.6	692	A34548	folliotropin recept
8	481.5	12.6	696	A41344	folliotropin recept
9	481	12.5	695	JN0898	folliotropin-chorlogon
10	478	12.5	695	JN0898	folliotropin recept
11	476.5	12.4	907	JE0176	orphan G protein-c
12	475.5	12.4	696	JC7361	folliotropin recept
13	471.5	12.3	700	A42395	luteinizing hormon
14	470.5	12.3	700	I77463	luteinizing hormon
15	470.5	12.3	700	A49744	luteinizing hormon
16	469	12.2	699	QRH0UT	luteinizing hormon
17	451.5	11.8	925	JC2033	G protein-coupled
18	447.5	11.7	764	I48882	thyrotropin recept
19	443	11.6	793	JC7390	thyrotropin recept
20	440.5	11.5	764	JC5643	thyroid stimulat
21	436	11.4	764	QRH0RH	thyrotropin recept
22	434	11.3	764	A35956	thyrotropin recept
23	430	11.2	764	A40077	thyrotropin recept
24	418	10.9	814	JC7389	thyroid stimulat
25	370	9.7	889	T20123	hypothetical prote
26	302	7.9	603	JC1282	insulin-like growt
27	296	7.7	603	A41915	insulin-like growt
28	294	7.7	603	JC6128	insulin-like growt
29	293.5	7.7	605	JC5239	insulin-like growt

30	286.5	7.5	1469	2	B36665	slit protein 2 pre
31	286.5	7.5	1480	2	A36665	slit protein 1 pre
32	285	7.4	1523	2	T13953	MEG's protein - ra
33	283.5	7.4	1531	2	T42218	slit-1 protein hom
34	283	7.4	1389	2	T13852	gene wheeler prote
35	281	7.3	961	2	T23395	hypothetical prote
36	280	7.3	1385	2	T13887	tlr protein - fru1
37	275.5	7.2	536	2	A34901	lysine carboxypept
38	273.5	7.1	1066	2	T15864	hypothetical prote
39	273	7.1	1091	2	A58532	glial cell membran
40	269.5	7.0	707	2	JC7763	neuronal leucine-r
41	268	7.0	361	2	A53660	chondroadherin pre
42	266.5	7.0	610	2	T23836	hypothetical prote
43	261	6.8	594	2	T23841	hypothetical prote
44	260	6.8	1051	2	T13174	gpi50 protein - fr
45	253.5	6.6	1134	1	A29944	chaoptin precursor

ALIGNMENTS

```
RESULT 1
S40241
G protein-coupled receptor - great pond snail
C:Species: Lymnaea stagnalis (great pond snail)
C>Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 15-Jun-2001
C:Accession: S40241
R:Tensen, C.P.; Kesteren, E.R.; Planta, R.J.; Cox, K.; Burke, J.F.; Heerikhuizen, H.;
submitted to the EMBL Data Library, June 1993
A:Description: A G protein-coupled receptor with LDL-binding motifs suggests a role f
A:Reference number: S40241
A:Accession: S40241
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1115 <TEN>
A:Cross-references: EMBL:Z23104; NID:9438128; PID:9438129
C:Superfamily: great pond snail LDL receptor-related G protein-coupled receptor; LDL
C:Keywords: G protein-coupled receptor; transmembrane protein
F:38-77/Domain: LDL receptor ligand-binding repeat homology <LDL1>
F:79-113/Domain: LDL receptor ligand-binding repeat homology <LDL2>
F:118-153/Domain: LDL receptor ligand-binding repeat homology <LDL3>
F:158-194/Domain: LDL receptor ligand-binding repeat homology <LDL4>
F:195-230/Domain: LDL receptor ligand-binding repeat homology <LDL5>
F:233-267/Domain: LDL receptor ligand-binding repeat homology <LDL6>
F:274-316/Domain: LDL receptor ligand-binding repeat homology <LDL7>
F:322-361/Domain: LDL receptor ligand-binding repeat homology <LDL8>
F:367-401/Domain: LDL receptor ligand-binding repeat homology <LDL9>
F:406-440/Domain: LDL receptor ligand-binding repeat homology <LDL10>
F:446-483/Domain: LDL receptor ligand-binding repeat homology <LDL11>
F:488-523/Domain: LDL receptor ligand-binding repeat homology <LDL12>
F:584-607/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR1>
F:608-631/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR2>
F:632-655/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR3>
F:656-679/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR4>
F:704-727/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR5>
F:774-797/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR6>
```

Query Match 21.7% Score 832; DB 2; Length 1115;
Best local similarity 30.0%; Pred. No. 7.7e-48;
Matches 205; Conservative 114; Mismatches 240; Indels 124; Gaps 13;

```

OY 27 CSLGFFPGGNTKCLPOLLHCNGVDDC-GNADADENCYV-----VICQC-----MSLPG 75
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 488 CLEQFRCRK-SFICNITKVCDDGVDCLOGMWDENCNVCYCPHQAICQCGVMTDCGQ 546
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 76 ELDMWKPTSVPSVSSNATSLQNLIRKLPPCFKRYHD-LQKLDQNNKITSISIA 134
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 547 KLEKEH---PYQWMEEDSKLMIGDNLNLSTFTFSATYKAYVLLDLSRHHLEIPYIS 602
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 135 FRGINSLSKLYLSNRRITFLKPGVEEDLHRLDWLIIEDNHLRSISPFTYGLNSLILVL 194
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 603 FQNNMKLTHLNLADNNITSLKNG-----LHNLHNLHNLHNLHNLHNLHNLHNLHNLH 625
```

```

QY 195 MNNVLRPLDKPLCOHMPRLHMLDLEGNHILMLRLFTISCSNLTVLVAKRKINHLNEN 254
Db 626 -----SLGLSLMKOLHNGKNIETIED 649
QY 255 TRAPLOKDELDGSKKIENTLPLPIKOLKELSQLMLSTNPLOKIOANOPDYIVKLSLS 314
Db 650 TESSMHLTVLDSNORLTHVYKNMFKGLOITVNLISROQINSINDGAFNNLANLYR 709
QY 315 LSGEISNTIOORFRLMLNLSHYEFKKFOYCGYAPHVRSCKPNTDGISSLENLASIIOR 374
Db 710 LSGNVTKDGQYKFMGLPRLVLEIKTDSYRCCILAPGVCSPKODFFSCEDLMSHHVLR 769
QY 375 VFWVVAVTCEGNIFVICMRPIRSEKNTLYAMSIISLCCADCMIGIYLFVIGDFLKER 434
Db 770 VSLVWGLVLAIVGNFVIEFMRVDRFGKSHFLITNLAIIGDLMGVYLLIINTADTYR 829
QY 435 GEYNKHAQIMESTHCQLVGSLAISTEVSVLLTFLTEKYICYIPRCVAPGCKRTI 494
Db 830 GYVISHDENMKOSGLCOFAGFVSTFSELSVLTLSITTDRLCILEPPLRRLGRLQAI 889
QY 495 TVLLIHWINGFIAPILPSNKEPFKNVYCGNVCPELHSDTESIGAQIYSVAIFLGINL 554
Db 890 IYMSCIHWLVFLAVLPDLGFTFEMFGRSGVCLALHVTDDRPGWE-YSGVFTILNL 948
QY 555 AAEIIVFSYSGMFYSV---HOSAITATEIRNOVKREMLAKRFFIYVTDALCWIPFV 611
Db 949 LSFVLVASSIYMF-SVAKKTRSAVTRAEBSKN---DNAMARMTLIVMDFCWPVITV 1003
QY 612 VKRLSLQVEIPGTTISWVYVIFLPIINSALNPILTYLTRPF-----KEKHIR 659
Db 1004 LGRVSLAGARADQVYMAVFLPLINSANPVIYTLSTAPLGNVYKRRANRFRKSFHS 1063
QY 660 F-----WYNYRORKS 669
Db 1064 FTGDTKHSYVDGTHSYCEKKS 1086

RESULT 2
JC1493
folliculin receptor - sheep
M:Alternate names: follicle stimulating hormone receptor
C:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)
C>Date: 03-Feb-1994 #sequence_revision 03-Feb-1994 #text_change 21-Jan-2000
C:Accession: JCI493; I47080
R:Khan, H.; Yarney, T.A.; Saitram, M.R.
Biochem. Biophys. Res. Commun. 190, 888-894, 1993
A:Title: Cloning of alternately spliced mRNA transcripts coding for variants of ovine
A:Reference number: JCI493; MUID:93176195
A:Accession: JCI493
A:Molecule type: mRNA
A:Residues: 1-695 <KHA>
A:Experimental source: testis
R:Yarney, T.A.; Saitram, M.R.; Khan, H.; Ravindranath, N.; Payne, S.; Seidah, N.G.
Mol. Cell. Endocrinol. 93, 219-226, 1993
A:Title: Molecular cloning and expression of the ovine testicular follicle stimulating
A:Reference number: I47080; MUID:93351750
A:Accession: I47080
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-695 <YAR>
A:CROSS-references: GB:L07302; NID:q165864; PIDN:AAA1525.1; PID:q165865
C:Genetics:
A:Gene: FSH-R
C:Superfamily: glycoprotein hormone receptor; leucine-rich alpha-2-glycoprotein repeat
C:Keywords: G protein-coupled receptor; glycoprotein; transmembrane protein
F:71-95/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR>
F:191,199/Binding site: carbohydrate (Asn) (covalent) #status Predicted

Query Match 13.1%; Score 501; DB 2; Length 695;
Best Local Similarity 24.4%; Pred. No. 6.5e-26;
Matches 168; Conservative 130; Mismatches 245; Indels 146; Gaps 28;

```

```

QY 40 CLPQLIHC-NGVDDCGNQADEDCVAVYLCCQMSLPGLLELDMMKPFYSVS-VSSNTAMS 97
Db 18 CHRLCHCSNG-----VFLCQDSKV-----TEPSSDLPRAVELR 52
QY 98 LQMWLRKLRLPPOCFKNYHDLOKLDLQNNKITSISIAFRGLNSLFTLYSHNRITFLKPG 157
Db 53 FVLTKLRLVPEGAFSGFGLDEKTEISQNDVLE-----IEAN 89
QY 158 VFEDRLRLMLIED-NHLSRISPPFTYGLNSLITLVNANNVTLRPLDKPLCOHMPRLHM 216
Db 90 VESNLPKLHEIRERKNNLLYIDPDAFQNLPLNRYLLISNTGKHLPAVHKIQSLQKV-L 148
QY 217 LDLEGN-HIHLNRLTEFISCSNLTVLV-MRKKNINHLNENTRAPLOKDELDGSK-KIE 273
Db 149 LDIODNINHTVERNSFMGLSFESMIWLSKNGIOEIHCAFCGTO-LDELNLSDMSNLE 207
QY 274 NLPPLIFKD-----LKEISQMLNSTP-----IQIOANOPDYIVKLSLS-LLEGIEI 320
Db 208 ELPNDVFQAGSPVILDISRTIRSLPSYGLENLKLAKKSTYHKKPLSLEKFTVLVEA 267
QY 321 S-----NIQOMFRLMNL-----H 336
Db 268 SLTYPSHCCAFAMWRQOTDLHPICNKSILROEVDMTQARGORISLADDEPSYAKGF 327
QY 337 IYFKKFOY--CGYAPHVRSCKPNTDGISSLENLASIIORVFWVVSAYTCGNIPIVGM 394
Db 328 MMYSEFDYDLCSEVDV-TCSPEDPAFNPCEDIMGDIIRVLWTFISLITAGNILLVLI 386
QY 395 RPIYRSEKNTLYAMSI--SLCCADCLMGYILFPIYGFEDLKFRGEYKHAQIMESTHCQL 452
Db 387 --LITSQYKLVPRFLMCAINFADLCIGYLLILASVDVHTKQYINVAIDMOTGACGA 444
QY 453 VGSIALISTEVSVLLTFLTEKYICYIPRCVAPGCKR---PIVLLIHWINGFI 507
Db 445 AGFTYFASBELSYTLFATLIERMHTITAMQL---ECKVHRHAASIMLVGWFAFV 500
QY 508 AFPIPLSNKEPFKNYGTNGVCPPLHSDTESIGAQIYVAIFGINLAFFIIVFSGSM 567
Db 501 ALFPI-----FGISSYMKVSGICLPM--DIDSPLSQLYVMSL-LVLNVLAFVYICGCTHI 552
QY 568 FYSVHOSAITATEIRNOVKREMLAKRFFIYVTDALCWIPVVFELSLOVEIGTIT 627
Db 553 YLTVRNPNITSS-----SSDTKIRAMAMLITTDLCNAPISFFAISLSKVPIL-TVS 605
QY 628 SWVVFIL--PINSALNPILTYLTRPF 654
Db 606 KSKILVLFVPIINSCANPFLYAITRNR 634

RESULT 3
I45896
follicle stimulating hormone receptor - bovine
C:Species: Bos primigenius taurus (cattle)
C>Date: 15-Oct-1996 #sequence_revision 15-Oct-1996 #text_change 21-Jan-2000
C:Accession: I45896
R:Ronde, A.; Lambert, A.; Saumande, J.; Silversides, D.W.; Lusier, J.G.
Mol. Reprod. Dev. 39, 127-135, 1994
A:Title: Structure of the bovine follicle-stimulating hormone receptor complementary
A:Reference number: I45896; MUID:95127199
A:Accession: I45896
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-695 <ROU>
A:CROSS-references: GB:L22319; NID:q404671; PIDN:AAC37324.1; PID:q404672
C:Genetics:
A:Gene: FSHR
C:Superfamily: glycoprotein hormone receptor; leucine-rich alpha-2-glycoprotein repea
F:71-95/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR>

Query Match 13.0%; Score 498; DB 2; Length 695;
Best Local Similarity 24.5%; Pred. No. 1e-25;

```



```

Db      541 EVVICGCIYHILYVFNPNIMSS-----SSDTKIAKRAMLIFPDLICMVPISFPAISA 594
QY      617 LLOVEIPGTTISWVYIFL--PINSALNPILYTLTTREPK 654
      1 : : : : : ||| | | : : : : :
Db      595 SLKVPFLI-TVSRKLTLVLFYRNPINCAMPFLYATFTKNFR 633
      1 : : : : : ||| | | : : : : :

RESULT      7
A34548
foliitropin receptor precursor - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 22-Jan-1993 #sequence_rev10n 22-Jan-1993 #text_change 13-Aug-1999
C:Accession: A34548; A41729
R:Spengler, R.; Braun, T.; Nikolics, K.; Segaloff, D.L.; Seeburg, P.H.
Mol. Endocrinol. 4, 525-530, 1990
A:Title: The testicular receptor for follicle stimulating hormone: structure and function
A:Reference number: A34548; MUID:91125358
A:Accession: A34548
A:Molecule type: mRNA
A:Residues: 1-692 <SEP>
A:Cross-references: GB:102842; NID:g204183; PIDN:AAA41175.1; PID:g204184
R:Heckert, L.L.; Daley, I.J.; Griswold, M.D.
Mol. Endocrinol. 6, 70-80, 1992
A:Title: Structural organization of the follicle-stimulating hormone receptor gene.
A:Reference number: A41729; MUID:92149579
A:Accession: A41729
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-692 <HEC>
A:Cross-references: GB:581198; NID:g245344; PIDN:AAB21415.1; PID:g245345
A:Note: sequence inconsistent with the nucleotide translation
R:Davis, D.; Liu, X.; Segaloff, D.L.
Mol. Endocrinol. 9, 159-170, 1995
A:Title: Identification of the sites of N-linked glycosylation on the follicle-stimulating hormone receptor
A:Reference number: A57562; MUID:95295729
A:Contents: annotation; glycosylation sites
A:Function:
A:Description: receptor that mediates the biochemical effects of follitropin
C:Superfamily: glycoprotein hormone receptor; leucine-rich alpha-2-glycoprotein repeat h
C:Keywords: alternative splicing; G protein-coupled receptor; glycoprotein; hormone rece
F:1-15/Domain: signal sequence #status predicted <SIG>
F:16-692/Product: follitropin receptor #status predicted <MAT>
F:16-692/Domain: extracellular hormone binding #status predicted <EHB>
F:56-70/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR3>
F:71-95/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR2>
F:96-120/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR3>
F:121-145/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR4>
F:146-169/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR5>
F:172-193/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR6>
F:194-218/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR7>
F:367-387/Domain: transmembrane #status predicted <TM2>
F:398-421/Domain: transmembrane #status predicted <TM2>
F:443-465/Domain: transmembrane #status predicted <TM3>
F:486-508/Domain: transmembrane #status predicted <TM4>
F:529-550/Domain: transmembrane #status predicted <TM5>
F:574-597/Domain: transmembrane #status predicted <TM6>
F:609-630/Domain: transmembrane #status predicted <TM7>
F:191,199,293/Binding site: cardiolipate (Asn) (covalent) #status predicted
F:534/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F:595/Binding site: phosphate (Ser) (covalent) (by protein kinase C) #status predicted

Query Match      12.6% Score 481.5; DB 2; Length 692;
Best Local Similarity 22.7% Pred. No.1.3e-24;
Matches 169; Conservative 140; Mismatches 272; Indels 165; Gaps 28;

QY      35 GNITKCLPQLHLCNGVDGCGNADENCVVVLCQCMSPGLGLDMKPFSTVPS-VSSNV 93
      1 : : : : : ||| | | : : : : :
Db      13 GTGSCGHHMLHCNS-----RVFLCQDSKY-----TELPDLPKNA 48
      1 : : : : : ||| | | : : : : :

QY      94 TAMSQWMLIRKRPDCEKFNHDLQKLDLQNNKTTTSIYAFRGINSLTKLKLSHNRITF 153
      1 : : : : : ||| | | : : : : :

```

Query Match 12.6% Score 481.5; DB 2; Length 696;
Best Local Similarity 23.3%; Pred. No. 1.3e-24;
Matches 144; Conservative 133; Mismatches 261; Indels 79; Gaps 17;

```

QY      87  PSVSNVYTAWSLQWNLIRKLPPDCFKNYHDLOKLDL-QNNKITSISIAFGRGINSILTKLY 145
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      45  PGRAGISRLSLTYFLPIKVIPOAFRGLENEVVKIEISQSLSEKIEANADNLINSEIL 104
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      146  LSNMR-ITFLAKPEVEDLHLEWLIIEDNLKISPT--FYGLNSILLYLVANNVTRL 202
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      105  IQMTKMLVYIEPAFNLPLPKYLSLCTNIGIRKLPDVTKIFSEFNILICNLHTTV 164
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      203  PDRPLCOHMRPLHMDLEGHINHLNLFISCSNLTVLVWRKNKIHLENTFAPLOKL 262
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      165  PAAFGGMNNESTITKLYGSGFEEIOSHANGNTLISLEKEMAHLLKMNDAFGRGRC 224
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      263  DELDLSGNKLENTPLPIFKDLKEL---SQLNSYNPLOKIOANOFYLVKLKSLSGIE 319
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      225  SILDISSTKQALPSYGLSIOFLIATSSYSLKKLPSEKREFTLLDATTLYPSHC--- 280
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      320  ISNIQGRMPRL---MNLSHIYEKKRQ----- 343
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      281  -----AFNLPPTKEONSFSTFKNFSCCESTARPPNETLYKALFAESELSDMDYD 333
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      344  -YGYAPHVNSCKPNTDGISSLNLASIIQRYFVWVWVAVTCFNIYICMPYIKSEN 402
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      334  GFC--SPKTLQCAPEDPAFNPCEIDIMGYDPLRVILINILINILAIMGNVTLVF--LLTSHY 389
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      403  KLTAMSLI--SLCCADCLMGIYFVIGFPLKRGEXNKAQIMVESTHCQLVGSALIS 460
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      390  KLVVPRFLMGNLSFADFECMLYLLLASVDAQYKGOYNNALIDMQNGCSVAGFTVFA 449
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      461  TEVSULLFLFLETKYICAIYPPRCVPRGCR--TIYVLLIMITGFVAPICPISNKEFK 519
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      450  SELSVTLTYITIERHTTYIAIQLODKLRLRAIPIMLGWLFSTILIAMPLVG----V 505
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      520  NNYGTNGVCPHLSSEDTESIGAQIYSVAIFLGINLAFTIIVESYSGSMFVSHQSAITAT 579
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      506  SSYMKVSICLPM---DVETLSQVYIILFI-LNVVAFIILICACIKIYPAVONPELMAT 561
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      580  EIRNQVKKEIILAKRFFIYFTDALCMIPFVVKFSLLOVEIPGTTISWVFIIL--PI 637
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      562  -----NKDKRIAKKAAVLLFTDTCMAPISFPAISAAKLVPLI-TVTNSKVLVLYEYV 614
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      638  NSALNPILYLTTRPPK 654
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      615  NSCANPFLVIAFTKAFR 631
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 9
JN0898
N:Alternate names: follicle stimulating hormone receptor (FSHR)
C:Species: Macaca fascicularis (crah-eating macaque)
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: JN0898; S36452
R:Gromoll, J.; Dankbar, B.; Sharma, R.S.; Nieschlag, E.
Biochem. Biophys. Res. Commun. 196, 1066-1072, 1993
A:Title: Molecular cloning of the testicular follicle stimulating hormone receptor of th
A:Reference number: JN0898; M0ID:94071854
A:Molecule type: mRNA
A:Accession: JN0898
A:Residues: 1-695 <CRO>
A:Cross-References: EMBL:X74454; NID:g336801; PIDN:CA52463.1; PID:g336802
A>Note: The authors translated the codon AAT for residue 488 as Arg
C:Function:
A:Description: receptor that mediates the biochemical effects of folliclin
C:Superfamily: glycoprotein hormone receptor; leucine-rich alpha-2-glycoprotein repeat
C:Keywords: G protein-coupled receptor; glycoprotein; hormone receptor; phosphoprotein;
F:1-11/Domain: signal sequence #status predicted <SIC>
F:18-695/Product: folliclin receptor #status predicted <PFI>
F:71-95/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR2>
F:96-120/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR3>
F:121-145/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR4>
F:146-166/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR5>
F:172-193/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR6>
F:194-218/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR7>

```

```

F:367-387/Domain: transmembrane #status predicted <TM1>
F:399-421/Domain: transmembrane #status predicted <TM2>
F:444-465/Domain: transmembrane #status predicted <TM3>
F:486-508/Domain: transmembrane #status predicted <TM4>
F:529-550/Domain: transmembrane #status predicted <TM5>
F:574-597/Domain: transmembrane #status predicted <TM6>
F:609-630/Domain: transmembrane #status predicted <TM7>
F:191,199,293,318/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:555/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status prediecte
F:596/Binding site: phosphate (Ser) (covalent) (by protein kinase C) #status prediecte

Query Match      12.5%; Score 481; DB 1; Length 695;
Best Local Similarity 24.8%; Pred. No. 1.4e-24;
Matches 158; Conservative 125; Mismatches 238; Indels 116; Gaps 25;

QY      99  QMWLIRKLPPDCFKNYHDLOKLDLQNNKITSISIAFGRGINSILTKLYLSNRI-TFLKPG 157
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      33  QESKVTIEPISDLPRNAIELR---FVHTKLVYIOGAFSGGDIKEKTEISONDVLEAYEAD 89
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      158  VFEDLHRLLEWLIETD-NHLSRISPTFYGINSILLYLVANNVLTRLPDKPLCOHMRPLH- 215
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      90  VFGNLPRLKHEIRIEKANNNLLIYNPEAFQNLPLRYLLISWTGIKHLPD-----VHKIHS 143
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      216  ---WLDLGN-HIHNRNLFTISC-NLTVLYMRKNKIHLENTFAPLOKDEIDLG- 268
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      144  FQKVLIDIDQINIHITERNSEFVGLSESVIWLNNKNGIOELINCAFNQTO-LDEINLSD 202
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      269  SNKIENPLPIFADKELSQLNSYNDPIOKIOANOFYLVKLKSLS-----LEGIE--- 319
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      203  NNNLELPNDVFHGASGPVLLDSRTRHSLPSYGLENLKLRKSTYNNLKLPSLEKLY 262
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      320  -----ISNIQGRM--FRPLMNL----- 335
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      263  ALMEASLTYPSSHOCAPANMRQISELPICKNSILHQEVDYMTQTRGQSSLAEDNESSY 322
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      336  ---HIYEKKFY--CGYAPHVNSCKPNTDGISSLNLASIIQRYFVWVWVAVTCFNI 389
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      323  SRGFDMTYAEFDYDLQNEVDV--TCSPKPDAPNPCEIDILGYNLIRVLIWFISTILATGNI 381
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      390  FVICMPRYISENKKLYAMSI-----SLCCADCLMGIYFVIGFPLKRGEXNKAQIM 444
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      382  IYL-----VLTTSQYKLVPRFLMGNLAFADLCIGIYLLILASVDLHRSQYHNVAIDW 436
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      445  MESTHCQLVGSALISTEVSULLFLETKYICAIYPPRCVPRGCR-----TIYVLL 499
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      437  QTGAGCDAGFFTFVPAFSELVYTLTAITLERMTITHAMQL-----DCKYVHRHAASVMW 492
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      500  IWITGFIVAFIPUSNKEFEKNYGTNGVCPHLSSEDTESIGAQIYSVAIFLGINLAFTI 559
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      493  GWIFAFAALFPI---FGISSYMKVSICLPM--DIDSPLSQLYVMSL-LVLNVLAFAV 544
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      560  IVSYSGSMFVSHQSAITATEINQVKKEIILAKRFFIYFTDALCMIPFVVKFSLLO 619
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      545  ICCCYTHIYLVNPNIVS-----SSDRIRAKRAMLFTFDLPMAISFPAISASLK 598
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      620  VEIPGTTISWVFIIL--PINSALNPILYLTTRPPK 654
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      599  VPLI-TVSAKAKILLVLYFPIINSCANPFLVIAFTKKNR 634
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 10
ORHUF
N:Alternate names: follicle stimulating hormone receptor (FSHR)
C:Species: Homo sapiens (man)
C>Date: 30-Sep-1991 #sequence_revision 06-Sep-1996 #text_change 22-Jun-1999
C:Accession: I57651; I56448; PC1147; S30560; I57672; JN0122
R:Gromoll, J.; Dankbar, B.; Gudermann, T.
Mol. Cell. Endocrinol. 102, 93-102, 1994
A:Title: Characterization of the 5' flanking region of the human follicle-stimulating
A:Reference number: I57651; M0ID:95011044

```

A:Accession: 157661
A:Status: translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-51 <GR0>
A:Cross-references: GB:S73199; NID:g685036; PIDN:AAB32071.1; PID:g685037
R:Gromoll, J.; Ried, T.; Holtgreve-Grez, H.; Nieschlag, E.; Gudermann, T.
J. Mol. Endocrinol. 12, 265-271, 1994
A>Title: Localization of the human FSH receptor to chromosome 2 p21 using a genomic probe
A:Reference number: I56448; MUID:95000244
A:Accession: 156448
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 286-695 <GR2>
A:Cross-references: GB:S73526; NID:g688069; PIDN:AAB32225.1; PID:g688070
R:Gromoll, J.; Gudermann, T.; Nieschlag, E.
Biochem. Biophys. Res. Commun. 188, 1077-1083, 1992
A>Title: Molecular cloning of a truncated isoform of the human follicle stimulating hormone
A:Reference number: PC1147; MUID:93075197
A:Accession: PC1147
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-223,286-294, 'P',296-342 <GR3>
A:Cross-references: EMBL:X68044; NID:g31473; PIDN:CAA48179.1; PID:g31474
R:Gromoll, J.
Experimental source: testis
submitted to the EMBL Data Library, August 1992
A:Reference number: S30560
A:Accession: S30560
A:Molecule type: mRNA
A:Residues: 1-12, 'R',14-223,286-294, 'P',296-342 <GR4>
A:Cross-references: EMBL:X68044; NID:g31473; PIDN:CAA48179.1; PID:g31474
R:Kelton, C.A.; Cheng, S.-Y.; Nugent, N.P.; Schweickhardt, R.L.; Rosenthal, J.L.; Overton,
Mol. Cell. Endocrinol. 89, 141-151, 1992
A>Title: The cloning of the human follicle stimulating hormone receptor and its expression
A:Reference number: 157672; MUID:93246012
A:Accession: 157672
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-679, 'N',681-695 <REL>
A:Cross-references: GB:SS9900; NID:g300072; PIDN:AAB26480.1; PID:g300073
R:Miyegishi, T.; Nakamura, K.; Takakura, Y.; Ibuki, Y.; Igataishi, M.
Biochem. Biophys. Res. Commun. 175, 1125-1130, 1991
A>Title: Cloning and sequencing of human FSH receptor cDNA.
A:Reference number: JN0122; MUID:91222171
A:Accession: JN0122
A:Molecule type: mRNA
A:Residues: 1-111,'T',113-196,'AV',199-306,'A',308-695 <MIN>
A:Cross-references: EMBL:M65085; NID:g182770; PIDN:AAA2477.1; PID:g182771
C:Genetics:
A:Gene: GDB:FSHR
A:Cross-references: GDB:I27510; OMIM:136435
A:Map position: 2p21-2p16
A:Introns: 223/3
A>Note: The exact position of the Intron cannot be determined from the experimental data
C:Function:
A:Description: receptor that mediates the biochemical effects of follitropin
C:Superfamily: glycoprotein hormone receptor; leucine-rich alpha-2-glycoprotein hormone repeat b
C:Keywords: alternative splicing; G protein-coupled receptor; glycoprotein hormone recept
F:1-695/Product: follitropin receptor precursor, long splice form #status predicted <SP1>
F:1-223,286-695/Product: follitropin receptor precursor, short splice form #status predi
F:1-125/Domin: signal sequence #status predicted <SIG>
F:1-223,286-695/Product: follitropin receptor precursor, short splice form #status predi
F:1-125/Domin: signal sequence #status predicted <SIG>
F:1-695/Product: follitropin receptor precursor, long splice form #status predicted <SP1>
F:1-223,286-695/Product: follitropin receptor precursor, short splice form #status predi
F:1-125/Domin: signal sequence #status predicted <SIG>
F:1-223,286-695/Product: follitropin receptor precursor, long splice form #status predicted <SP1>
F:1-223,286-695/Product: follitropin receptor precursor, short splice form #status predi
F:1-125/Domin: signal sequence #status predicted <SIG>
F:1-223,286-695/Product: follitropin receptor precursor, long splice form #status predicted <SP1>
F:1-223,286-695/Product: follitropin receptor precursor, short splice form #status predi
F:1-125/Domin: signal sequence #status predicted <SIG>
F:1-223,286-695/Product: follitropin receptor precursor, long splice form #status predicted <SP1>
F:1-223,286-695/Product: follitropin receptor precursor, short splice form #status predi
F:1-125/Domin: signal sequence #status predicted <SIG>
F:1-223,286-695/Product: follitropin receptor precursor, long splice form #status predicted <SP1>
F:1-223,286-695/Product: follitropin receptor precursor, short splice form #status predi
F:1-125/Domin: signal sequence #status predicted <SIG>
F:1-223,286-695/Product: follitropin receptor precursor, long splice form #status predicted <SP1>
F:1-223,286-695/Product: follitropin receptor precursor, short splice form #status predi
F:1-125/Domin: signal sequence #status predicted <SIG>
F:1-223,286-695/Product: follitropin receptor precursor, long splice form #status predicted <SP1>
F:1-223,286-695/Product: follitropin receptor precursor, short splice form #status predi
F:1-125/Domin: signal sequence #status predicted <SIG>
F:1-223,286-695/Product: follitropin receptor precursor, long splice form #status predicted <SP1>
F:1-223,286-695/Product: follitropin receptor precursor, short splice form #status predi
F:1-125/Domin: signal sequence #status predicted <SIG>
F:1-223,286-695/Product: follitropin receptor precursor, long splice form #status predicted <SP1>
F:1-223,286-695/Product: follitropin receptor precursor, short splice form #status predi
F:1-125/Domin: signal sequence #status predicted <SIG>
F:1-223,286-695/Product: follitropin receptor precursor, long splice form #status predicted <SP1>
F:1-223,286-695/Product: follitropin receptor precursor, short splice form #status predi
F:1-125/Domin: signal sequence #status predicted <SIG>
F:1-223,286-695/Product: follitropin receptor precursor, long splice form #status predicted <SP1>
F:1-223,286-695/Product: follitropin receptor precursor, short splice form #status predi
F:1-125/Domin: signal sequence #status predicted <SIG>
F:1-223,286-695/Product: follitropin receptor precursor, long splice form #status predicted <SP1>
F:1-223,286-695/Product: follitropin receptor precursor, short splice form #status predi
F:1-125/Domin: signal sequence #status predicted <SIG>
F:1-223,286-695/Product: follitropin receptor precursor, long splice form #status predicted <SP1>
F:1-223,286-695/Product: follitropin receptor precursor, short splice form #status predi
F:1-125/Domin: signal sequence #status predicted <SIG>
F:1-223,286-695/Product: follitropin receptor precursor, long splice form #status predicted <SP1>
F:1-223,286-695/Product: follitropin receptor precursor, short splice form #status predi
F:1-125/Domin: signal sequence #status predicted <SIG>
F:1-223,286-695/Product: follitropin receptor precursor, long splice form #status predicted <SP1>
F:1-223,286-695/Product: follitropin receptor precursor, short splice form #status predi
F:1-125/Domin: signal sequence #status predicted <SIG>
F:1-223,286-695/Product: follitropin receptor precursor, long splice form #status predicted <SP1>
F:1-223,286-695/Product: follitropin receptor precursor, short splice form #status predi
F:1-125/Domin: signal sequence #status predicted <SIG>
F:1-223,286-695/Product: follitropin receptor precursor, long splice form #status predicted <SP1>
F:1-223,286-695/Product: follitropin receptor precursor, short splice form #status predi
F:1-125/Domin: signal sequence #status predicted <SIG>
F:1-223,286-695/Product: follitropin receptor precursor, long splice form #status predicted <SP1>
F:1-223,286-695/Product: follitropin receptor precursor, short splice form #status predi
F:1-125/Domin: signal sequence #status predicted <SIG>
F:1-223,286-695/Product: follitropin receptor precursor, long splice form #status predicted <SP1>
F:1-223,286-695/Product: follitropin receptor precursor, short splice form #status predi
F:1-125/Domin: signal sequence #status predicted <SIG>
F:1-223,286-695/Product: follitropin receptor precursor, long splice form #status predicted <SP1>
F:1-223,286-695/Product: follitropin receptor precursor, short splice form #status predi
F:1-125/Domin: signal sequence #status predicted <SIG>
F:1-223,286-695/Product: follitropin receptor precursor, long splice form #status predicted <SP1>
F:1-223,286-695/Product: follitropin receptor precursor, short splice form #status predi
F:1-125/Domin: signal sequence #status predicted <SIG>
F:1-223,286-695/Product: follitropin receptor precursor, long splice form #status predicted <SP1>
F:1-223,286-695/Product: follitropin receptor precursor, short splice form #status predi
F:1-125/Domin: signal sequence #status predicted <SIG>
F:1-223,286-695/Product: follitropin receptor precursor, long splice form #status predicted <SP1>
F:1-223,286-695/Product: follitropin receptor precursor, short splice form #status predi
F:1-125/Domin: signal sequence #status predicted <SIG>
F:1-223,286-695/Product: follitropin receptor precursor, long splice form #status predicted <SP1>
F:1-223,286-695/Product: follitropin receptor precursor, short splice form #status predi
F:1-125/Domin: signal sequence #status predicted <SIG>
F:1-223,286-695/Product: follitropin receptor precursor, long splice form #status predicted <SP1>
F:1-223,286-695/Product: follitropin receptor precursor, short splice form #status predi
F:1-125/Domin: signal sequence #status predicted <SIG>
F:1-223,286-695/Product: follitropin receptor precursor, long splice form #status predicted <SP1>
F:1-223,286-695/Product: follitropin receptor precursor, short splice form #status predi
F:1-125/Domin: signal sequence #status predicted <SIG>
F:1-223,286-695/Product: follitropin receptor precursor, long splice form #status predicted <SP1>
F:1-223,286-695/Product: follitropin receptor precursor, short splice form #status predi
F:1-125/Domin: signal sequence #status predicted <SIG>
F:1-223,286-695/Product: follitropin receptor precursor, long splice form #status predicted <SP1>
F:1-223,286-695/Product: follitropin receptor precursor, short splice form #status predi
F:1-12

F:486-508/Domain: transmembrane #status predicted <TM>
F:529-550/Domain: transmembrane #status predicted <TM>
F:574-597/Domain: transmembrane #status predicted <TM>
F:609-630/Domain: transmembrane #status predicted <TM>
F:191,199,293,318/Binding site: carbohydrate (asn) (covalent) #status predicted
F:555/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F:596/Binding site: phosphate (Ser) (covalent) (by protein kinase C) #status predicted

Query Match 12.5%: Score 478; DB 1; Length 695;
Best Local Similarity 23.1%: Pred. No. 2,26-24;
Matches 160; Conservative 129; Mismatches 250; Indels 154; Gaps 26;

QY	40	DB	18	QY	99	DB	54	QY	159	DB	91	QY	216	DB	145	QY	270	DB	204	QY	320	DB	264	QY	336	DB	324	QY	391	DB	383	QY	449	DB	441	QY	504	DB	497	QY	564	DB	549	QY	624	DB	603
CLPQLHNGVDGDCQNADEDCVVLCCQCSLPLGELDMKKPFTSVPS-VSSNVTASL	18	CHHRCHCSN-----RVFLCQ-----ESKYTEIPSDLPRAHLEIRF	99	QWNLIRKLPRDCFKNYHDQLQKLDLQNNKRTISITVAFRGLNSLTKLYLSHNRITFLKPGV	54	VLTFRVIOKGAFSGFDGLKEIKESQNDVLEY-----LEADV	159	FEDLHRLRLMLIED-NHLSRISPPFTGYLSLLVLVANNVLTFLRPLQOHMRLH--	91	FSNLPKLEIRIREKANNLLTYINPEAFQNLPLQYLLISNTGIRKLDP-----VAKIHS	216	WDLDEGN-HIHRLRLTFSCS-NLTVMYMRKNINHLENMFAPQLKDELIDG-S	145	QKVLLDIDQDNIHNIERNSPYGLSFESVILMLKNGIOELHNCAPNTQ-LDELNSDN	270	NKIENLPPLIFKDKELSQLNSLYNPQKIQANQFDYLVKLSISLEGIE-----	204	NMLELPMDVFGASGPYLLDSIRTRISLSPSGLENLKLIRASTYMLKKLPLEKLV	320	-----ISNIQQRK-FRPLMANS-----	264	LMESALTYPSHCCAFANMRQISLHPLICNKSILNQEVDMTQTRGRRSSLAEDNESSYS	336	---HYFFKFOY--CGYAPHYRSCKPNFDGSSLENLASSIIQRYVFWVWSAVTCGNIF	324	RGFDMTYIEFDYDLCEYVDV-TCSPPKPARPCPDINGVNLKRLIMFISILATIGNII	391	VICMRPYRSENKILYAMSI--SLCCADCLMGIVLDFVIGGFDLKERGEYKNHAQIMEST	383	VLVI--LFTSQYKLIVPRFLMCMNLAFAADLCIGIYLLLASVDIHKSGYHNVAIDMGGA	449	HCQLVGSLAISTEYVULLFLFLTEKTYICIVYPRCVRPKCR-----TITVLILMIT	441	GDAAGFTVFSELSVYTLTAITLERMHTITHAMQ-----DCKQLRHAASVVMGWIF	504	GRIVAFPLSNKEPFKNYGTGVCYCPHLSSEDTESIGAOIVSAVLEGINLAAFIIYFS	497	AFMAALPL-----FQISSYMKVSTCLPM---DIDPSLSQLVMSL-LVLNVLAFVITGCG	564	YGSMEYSVHQSAITATEIRNOYKREMLAKRFFETVFDALQWIPFVYKELSLQVEIP	549	YHIIYLVARNPIVSS-----SSDTRIAKRRAMLIFDFCLCMAPISFPFASIASLKVPLI	624	GTITSMVYFIL--PINSALNPILTYTLTRPRK	603	TVSKAKILLVLEHPINSCANPELYAIFTKMR	

RESULTS
JB0176
orphan G protein-coupled receptor precursor - human
C:Species: Homo sapiens (man)
C:Date: 03-Jul-1998 #sequence_revision 10-Jul-1998 #text_change 21-Jul-2000
C:Accession: JE0176
R:McDonald, T.; Wang, R.; Bailey, W.; Xie, G.; Chen, F.; Caskey, C.T.; Liu, Q.
A:Title: Identification and Cloning of an orphan G protein-coupled receptor of the g11
A:Reference number: JE0176; MUID:98308104
A:Accession: JE0176

A:Molecule type: mRNA
A:Residues: 1-907 <MCD>
A:Cross-references: GB:AF062006; NID:g3366801; PIDN:ACC28019.1; PID:g3366802
C:Comment: This protein is a receptor for a novel class of glycoprotein ligands.
C:Genetics:
A:Gene: HG38
A:Map position: 12q22-23
F:1-21/Domain: signal sequence #status predicted <SIG>
F:562-583/Domain: transmembrane #status predicted <TM1>
F:594-616/Domain: transmembrane #status predicted <TM2>
F:639-660/Domain: transmembrane #status predicted <TM3>
F:681-701/Domain: transmembrane #status predicted <TM4>
F:725-744/Domain: transmembrane #status predicted <TM5>
F:768-791/Domain: transmembrane #status predicted <TM6>
F:803-824/Domain: transmembrane #status predicted <TM7>

Query Match 12.4%; Score 476.5; DB 2: Length 907;
Best Local Similarity 23.0%; Pred. No. 3.9e-24;
Matches 195; Conservative 134; Mismatches 316; Indels 203; Gaps 28;

```
QY 28 SLGYFPCGNIT-----KCLPOLHCGVDCGQADENCVVVLCQCMSPGLIEDMKRP 82
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 101 ALTYIPKGFGLGSLKVL--MLQNNQLHVPPEA-----LQVLRSLQSLRDANI 150
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 83 FTSVPSVSSNVTAMSLQW--NLIRRLPPDCFKNHDQKLDLQNNKITSISYAFRGIN 139
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 151 SYVPPSCFGLSHLRMLMDNALTEIPQAFRSLAQAMTLALNKIHIDPYAGNLS 210
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 140 SLTKYLISNNRTFLKPGFEDLHRLLEMLIEDNHSRISPPFYGLNSLILVLMNNVL 199
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 211 SLVVLHLNHNRLHSIGKCGDGLHSLLETLDLVNNDLDEF-PTAIRLSNMLKELGFHSNNI 269
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 200 TRLPPKPLC-----OHMRLHMLDLEG-NHITHNLNLFISG 235
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 270 RSIPPEAFPNQNSLTIHFHYDNPIDQVGSAPQHLPBELFTLTLNGASQITTEPDLT--GT 327
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 236 SNLTVLVMKRNKINHLNENTFAPLQKLDLDELGSNKIENLPILFIDKLELSQNLSTYRP 295
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 328 ANLESITLGAQISSLPQYVNCQPLQVLDLSYNLELDPS--FSVCQKIQKIDLRHNE 385
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 296 IOKIQANODIYLVKLSLSLEGIEINIOQRMRLPMN-----L 334
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 386 IYEIKVDYTFQQLSLKSLMLAMNKIAIHPNASTPLSLIKLDLSNLLSPFTGLHGI 445
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 335 SHIYFK-----KFOYCGY----- 347
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 446 THIKLTLGNALQSLISSENPPELKYTEMRYAQCCAFVGCENAKYKTSNQMNKGDSNMD 505
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 348 -----APHYRSCKPNITDGIISLENILASTIIOREYFW 378
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 506 LHKKDGAMFOADERDELDLDFEEDLRKALHSVQCSPPSGPRKPECHLDLGMIRIGYW 565
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 379 VVS--AVTQPGNITFV---ICMRYISENKKIYAMSTISLCCADCLMGIFYLVIGSDKLF 433
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 566 TIAVALTLC--NALVSTYFRSPLYSPIKLL--IGVIAAVMMLGVSSAVLAGVADFT 620
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 434 RGEYNKHAOLMESTRCOLVGLSLAISTEVSLLTFLEKTYICIVYFRVCRPCKRT 493
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 621 PGSFARHGAMWENGVCYHIGLSTIFASSSVFLTLALALERFSYKAKFTKAPFS 680
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 494 ITVLILIM-ITGFIYAFIPLSNKEFFKNYYGNGVCFPLHSEDTESIGQIYSVALFLGI 552
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 681 LKVIITLCLALLTMAAVPLLG-----GSKYGASPLCLPIPGEPSTMG--YVVALIL-L 732
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 553 NLAAPFIIVFSVGSMPYSHVQSAITATEIRNOVKKEMIIAKRFFIYFDALCWIPIFV 612
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 733 NSLCFMMMTIATKLYCNLDKG-----DLENIWDCSMV--KATALLFNFCLINCVAFL 785
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 613 KFLSLQVE-IPQITSWVFIPLINSALNPILYTLTRPFKEMIHFR-----WYNTR 665
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 786 SFSSILNLFPISEVYIKFILLVVPPLACINPLLITLFPNHFREDLVSLRKQYVWTRSK 845
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```
QY 666 -----QKRSNDS-KGQKTYAPSFIVEMXPLQEMPELMKPDLTFFPCMSLI 712
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 846 HPSLMSINSDVEKOSCDSTQALVFTSSITYDLPSSVSP-----ATP-----V 892
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 713 SOSTRLNS 720
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 893 TESCHLSS 900
```

RESULT 12

JC7361

folitropin receptor precursor - newt
M:Alternate names: follicle-stimulating hormone receptor

C:Species: Cynops pyrrhogaster (newt)

C:Date: 17-Nov-2000 #sequence_revision 17-Nov-2000 #text_change 31-Dec-2000

C:Accession: JC7361

R:Nakayama, Y.; Yamamoto, T.; Oba, Y.; Nagahama, Y.; Abe, S.

Biochem. Biophys. Res. Commun. 275, 121-128, 2000

A:Title: Molecular cloning, functional characterization, and gene expression of a fo

A:Reference number: JC7361

A:Contents: Testis

A:Accession: JC7361

A:Molecule type: mRNA

A:Residues: 1-696 <NAK>

A:Cross-references: DBJ:AB005587

C:Comment: This protein, containing seven transmembrane domains and a large glycosyla

in and thyroid stimulating hormone. This receptor has a common signal transduction pa

C:Genetics:

A:Gene: fsh-r

C:Superfamily: glycoprotein hormone receptor; leucine-rich alpha-2-glycoprotein resea

C:Keywords: disulfide bond; glycolysis; glycoprotein; hormone receptor; testis; trans

F:1-11/Domain: signal sequence #status predicted <SIG>

F:18-696/Product: follicle-stimulating hormone receptor #status predicted <MAT>

F:370-389/Domain: transmembrane #status predicted <TM1>

F:402-424/Domain: transmembrane #status predicted <TM2>

F:447-468/Domain: transmembrane #status predicted <TM3>

F:489-511/Domain: transmembrane #status predicted <TM4>

F:532-553/Domain: transmembrane #status predicted <TM5>

F:577-600/Domain: transmembrane #status predicted <TM6>

F:612-633/Domain: transmembrane #status predicted <TM7>

F:465-190,198,267,292/Binding site: carbohydrate (Asn) (covalent) #status predicted

F:445-520/Disulfide bonds: #status predicted

Query Match 12.4%; Score 475.5; DB 2: Length 696;
Best Local Similarity 23.4%; Pred. No. 3.3e-24;
Matches 163; Conservative 135; Mismatches 250; Indels 149; Gaps 30;

```
QY 32 FPGCGNTKCLPOLHCGVDCGQADENCYVVLCCQCMSPGLIEDMKKPTFSVP-SVS 90
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 16 FGHYPVCRCLINRYFT-----QESHVV-----QIPRDP 44
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 91 SNTVAMSLOMNLIRKLPDPCFNHYHDQKLDL-QNNKITSISYAFRGINSITKLTLSH 148
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 45 KNSTELREVLTKYTVYLPKAAFSGFEDVENIETISQNDVLTTEANVNSHLPKLREIIEKA 104
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 149 NRTFLKPGVFEDLHRLLEMLIEDNHSRIISPTFYGLNSLILVLMNNVFLRLDPKPLC 208
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 105 NMNVYIDPAPQNLPLSKYLLISNTGIQV--PAVSKIRS-----FHSVL 147
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 209 QHMPRLHMLDLEGN-HIHMDRLNITFISGNSLYLV-MRKNKINHLNENTFAPLQKIDEID 266
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 148 -----LDVODNINIRHIGKNSFAGLSSESITIRLNNGIEIQNHAENGTH-LNEIN 198
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 267 LGSN-KIENLPPLIFKD-----LKELSQNLISYNP-----IOKIOANDFDLYVKLSLS 314
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 199 LSDNQRLKLPDQVFGAGNPVILDISKRIRIHLPLPNNGIENIKKFAARNYIYIKLPIPE 258
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 315 --LEGIE-----ISNIOQRM--FRPLMN-----L 334
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 259 KPAELLEANLTYPSHCCAPANRERKKSEMHPICNKSFSGHDAEKPEKDNLRFSNEDYL 318
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```


Query 335 SHIYK-----KFQY----CGAYPHVRSCKRNTGDISSELMULASTIORVWVYSA 382
| :
Db 319 SSYGFSYSLVENGDGFNDYLILCNEV--HDVICPKRPAPFNCPEIDMGDNFLRWIMWTISI 377
VTCCGNIVICMRPYIRENKLTYAMSTII--SLCADCLMGLIEYVIGGSFDLKEFGENTKH 440
363 : | :
Db 378 LATGNITVLVI--LISSQYFTYPRFLMCNLARADLCMGYILLLLASVDTKTSQYYNH 435
441 AQLMMSEHCOLVSGLSALSTEVSALLTFPLTEKYICIVPFCBVGRCRTYTVLILI 500
Db 436 ALDMQTGGSCAAAGFFTYFASELSTYLTFTTIERMHIITYAMOLDKVRERRHAFTALMV 495
501 -WTGEFYAIPDLSKKEPFKNYGTNGVCPDLHSDTESIGAQIYSAIFGLINLAFTI 559
Db 496 GWIFAFTVALPIPI---FGVSSYTKVSIICLP---DIERSQSQSYIFEI-LVLNVCAFLI 547
IVESGSMFEYSHGSAITATEIRNOVKKEMTLAKRFETIVTDALCMPIPVAFELSLQ 619
Db 548 ICACYTGILYIVRNPNVSS-----NSDTKARKMALITFDPLCAMPISFPAISASLK 601
VEIPGTISWVFIPL--PINSALNPILYTTTPRK 654
Db 602 IPLI-TVESKILLYLFYPINSCANPFLYAIFTXTER 637

RESULT 13
A42395
luteotropin receptor - mouse
N:Alternate names: lutealizing hormone-choriogonadotropin receptor
C:Species: Mus musculus (house mouse)
C>Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 13-Aug-1999
C:Accession: A42395
R:Guidermann, T.; Birnbaumer, M.; Birnbaumer, L.
J. Biol. Chem. 267, 4479-4488, 1992
A>Title: Evidence for dual coupling of the murine lutealizing hormone receptor to adenylyl
monoreceptor expressed in L cells
A:Reference number: A42395; MUID:92165799
A:Accession: A42395
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-700 <GDU>
A:Cross-references: GB:W81310; GB:W81318; NID:g198811; PIDN:AAA9437.1; PID:g198812
A>Note: Sequence extracted from NCBI Backbone (NCBIT:4064, NCBI:P:4066)
C:Superfamily: glycoprotein hormone receptor; leucine-rich alpha-2-glycoprotein repeat h
C:Keywords: G protein-coupled receptor; transmembrane protein
E:54-77/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR1>
E:78-102/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR2>
E:103-117/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR3>
E:128-152/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR4>
E:153-177/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR5>
E:180-201/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR6>
E:202-226/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR7>

Query Match 12.3%, Score 471.5; DB 2; Length 700;
Best Local Similarity 23.8%; Pred. No. 6.2e-24;
Matches 148; Conservative 131; Mismatches 271; Indels 71; Gaps 18;

Db 87 PSVSNVTAMSLOMLIKRLPDPCFKNYHQDKIDL--ONNKITSISIAFGINSTIKLY 145
| :
Db 49 PEPRIGLARLSTLYPVKVIPSQAARGINEVKKLEISDSLSIEREANAFNNLNLSEIL 108
LSHNK--ITFLKGVGEDHLRLEMLITEFN-----HLSRISPFPFYGNSLILVLMNNV 198
109 IONTNKLXIIEGATNTNPLRKLYSICNTGIRTLDPVKISSSEP-----NFLIECDMLY 164
199 LTRLDKPLCOHMPRLHWLDEGNHHNLRLMTFISCNSLTVLYVRKKNI--HLENEN 254
165 ITTIGNAFPOGMNNSITFLKLYGNGEEVGSHAF---NGTTLISLEKENILEYKMHSG 220
255 TPAPLQKDELDDGSKNKIENLP-----PLTFDKLELSOL-----NLST- 293
221 TFQATGTSILDVSSCTQALPSHGLESIQTLTIASSYSLTVLPREKFTSLVATTVP 280

```

OY 294 -----NPiOKiQANQDYLvTKLKSLSBEiEiSNIQQRFRiLMLSHiYFKKPYCG 346
Db 281 SHCCAFRLPKRKEQNFSEiFENFSKOCESiTBREAN-NETLYSAiFEBENELSGMDYDYDF 339
OY 347 YAPHRSCKPMPTDGGiSSiENLSTiORiEVPVWVAVSAYTCGNiVICMRPiRiSENKLYA 406
Db 340 CSPKTLQCTPEPBDAFNPCEDiMGYAFiRLVLiWLiNLiAIGNLiVLiVLLTSRKLiLVPiR 399
OY 407 MSiISLiCCADLiMGiYLiFViGPFiLKEFEGNKHQAOLMESTHCQLiVSLiSTEVSVL 466
Db 400 FLiMGNLiSPADFCMGiYLiLLiASVDSQFKQYiNNAiDMQSGSCSAiGFFiVFAiSELiSV 459
OY 467 LiTFiTEKiYCiYiPFiFCVPRiGKCR-TiYVLiLiLiWiTGFViAFiPiPSiNKEFENYiGTN 525
Db 460 TiVTiTLiTRWHTiYiYAVOLiDQKLiRLiRAHiPiPLiMGiWFiSTLiMAVLiPLVG-----VSSiTKV 515
OY 526 GVCiFPLiHEDESiEGiAQiYSAiFiGLiNLAiFiTiVSiSGSiMFYSiQSAiTiTATEiRiNOV 585
Db 516 SiCLiPM-----DVESiTLiSiYLiLiSiLL-LiNAVAFViCiACVYRiYiFVAVONPDELTAiP----- 565
OY 586 KEMiLNAiRFFiYiFETDALCiMPLiFVVKLiSLLOVEiLPGiTiTSWViFiLiL-PiNSLiNP 643
Db 566 NKDTRiAKKMAiLiLFTDFTCMAiPiSFaiSAiFVPLi-TViTNSKVLiVLiFYPVNSCANP 624
OY 644 LiYTLiTRPEKE---MiHRF 660
Db 625 FiYAVTiFKAiFORDFiLLiSRF 645

RESULT 14
177463
Inteinizing hormone/chorionic gonadotropin receptor - rat
C.Species: Rattus sp. (rat)
C.Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 21-Jan-2000
C.Accession: I77463
R.Atsinikl, J.T.; Pietila, E.M.; Iakkakorpi, J.T.; Rajanleml, H.J.
MOL. Cell. Endocrinol. 84, 127-135, 1992
A.Title: Expression of the LH/CG receptor gene in rat ovarian tissue is regulated by
A.Reference number: 157668; MUID: 92347604
A.Accession: I77463
A.Status: preliminary; translated from GB/EMBL/DBJ
A.Residues: 1-700 <RES>
A.Molecule type: DNA
A.Cross-references: GB:540803; NID:g252163; PION:AMB22680.1; PID:g252164
C.Genetics:
A.Introns: 58/2; 82/2; 107/2; 132/2; 157/2; 183/2; 206/2; 231/2; 293/2; 320/2
C.Superfamily: glycoprotein hormone receptor; leucine-rich alpha-2-glycoprotein repea
F:153-177/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR>

Query Match 12.3%; Score 470.5; DB 2; Length 700;
Best Local Similarity 22.8%; Pred. No. 7.2e-24;
Matches 149; Conservative 135; Mismatches 233; Indels 137; Gaps 23;

OY 87 PSVSSNVTAMSiQNMiLiRKLPPDCFKNYHDiLOKLiDL-ONNKiTSiSiYAFRGLNSLiTKLY 145
Db 49 PGRPRGLARLiSTiYLiPVAViPSQAFRGiLNEYVKiEiSODSLERiEANAiFNLiLSELiL 108
OY 146 iSHNR-TiFFLiAFGVELiLHRELiEiLiEDN-----HLiRiSPiPFFYGLNSiLiLLVLMNNV 198
Db 109 iONTKLiLiYiEiGAFiTNiPLRiKYLiSiCWTGiRTiLTPDYKiSSiSEF-----NFiLiLiCNLiH 164
OY 199 LiRLiDLiPCLiCQiMPLiRLiHMLiDEGNiHNLiRNLiTFiSCNLiLVLiWKRKKiNHLiMENiFAP 258
Db 165 iTTiFGNAiFQGNNSiSVTLiKLiYNGiEEVQSHAF---NGTiLiLiSLiL-----LKENiYi- 213
OY 259 iOKLiDE-----LiDGiSKNKiEiNLPiLiFKDLiKELiSOLiNSiYNPiOKiQANQDYL 307
Db 214 LKMiNSGAiFQAGiTGSiLiDLiSTKiQALPS-----HLESiQF-----Li 252
OY 308 VLiKLSLiEGLiEiSNIQQRFRiPLM-----NLSiYiFKKPiQF--- 343

```

Db 253 IALSYSLKTLTP-----SKEKFTSLVATLTTPSHCAFRMLPKKEQNFSPSIFENFSKOC 308
 A:Accession: A32460
 A:Molecule type: protein
 A:Residues: 27-32, 'X', 35-37 <ROC>
 C:Genetics:
 A:Insertions: 58/2, 82/2, 107/2, 132/2, 157/2, 183/2, 206/2, 231/2, 293/2, 320/2
 C:Superfamily: glycoprotein hormone receptor; leucine-rich alpha-2-glycoprotein repeat
 C:Keywords: alternative splicing; G protein-coupled receptor; glycoprotein; hormone r
 Db 309 ESTVRRADNETLYSAIFEENELSGWDYDGC--SPKTLQCAPPPDAFNPCEDDMGYAF 366
 QY 374 RVEVWVAVATCGNIFVICMRPYINSEKLYAMSIISLCCADCLMGIFYVIGFPLKF 433
 F:54-77/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR1>
 Db 367 RVLIMINILAIIGNLVLFVLLTSRYKLVPRLMCLNSFPADCMGLYLLILASVDSQT 426
 QY 434 RGEYKHAQIAMESTHCOLVGSALISTEYVLLFTLEKYCIYPRPCVPGKCR- 492
 F:103-127/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR3>
 F:128-152/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR4>
 F:153-177/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR5>
 F:180-201/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR6>
 Db 427 KGOYNNHAIWMQSGGCGAAGFFVFASELVTLVTLTLEKMTITTYAVOLDKLRLH 486
 QY 493 TITVLLIMITGFIATFIPLSNKEFFKNYGTGVCPLSEDESTIGAOIYSAIFLGI 552
 F:202-226/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR7>
 F:268-389/Domain: transmembrane #status predicted <TM1>
 F:400-422/Domain: transmembrane #status predicted <TM2>
 F:444-466/Domain: transmembrane #status predicted <TM3>
 F:489-511/Domain: transmembrane #status predicted <TM4>
 F:530-551/Domain: transmembrane #status predicted <TM5>
 F:575-598/Domain: transmembrane #status predicted <TM6>
 F:610-631/Domain: transmembrane #status predicted <TM7>
 F:603-178, 199, 295, 303, 317/Binding site: carbohydrate (Asn) (covalent) #status predict
 F:647, 648/Binding site: palmitate (Cys) (covalent) #status predicted
 F:681/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicte
 Db 593 AISAARVPLI-TVTNKILLVLFYVNSCANPFLYAIPTKAFORDFLLLSRF 645
 RESULT 15
 A49744
 lutropin-choriogonadotropin receptor precursor - rat
 N:Alternate names: lutelizing hormone-chorogonadotropin receptor
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 03-May-1994 #sequence, revision 13-Sep-1998 #text_change 13-Aug-1999
 C:Accession: A49744; A40545; A41343; A61453; A32460
 R:KOC, Y.B.; Ji, T.; Slaughter, R.G.; Ji, T.H.
 E:Endocrinology 120, 2297-2308, 1991
 A:Title: Structure of the lutelizing hormone receptor gene and multiple exons of the co
 A:Reference number: A49744; MOID:91209270
 A:Accession: A49744
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-700 <KOC>
 A:Cross-references: GB:M68917
 A:Note: authors translated the codon CAA for residue 307 as Glu, AAC for residue 355 as
 R:Tsai-Morris, C.H.; Buczko, E.; Wang, W.; Xie, X.Z.; Dufau, M.L.
 J. Biol. Chem. 266, 11355-11359, 1991
 A:Title: Structural organization of the rat lutelizing hormone (LH) receptor gene.
 A:Reference number: A40545; MOID:91250455
 A:Accession: A40545
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-320 <RSA>
 A:Cross-references: GB:M63918; GB:M63919; GB:M63920; GB:M63921; GB:M63923; GB
 R:McFarland, K.C.; Sprengel, R.; Phillips, H.S.; Koehler, M.; Rosembilt, N.; Nikolics, K
 Science 245, 494-499, 1989
 A:Title: Lutropin-choriogonadotropin receptor: an unusual member of the G protein-couple
 A:Reference number: A41343; MOID:89332512
 A:Accession: A41343
 A:Molecule type: mRNA
 A:Residues: 1-700 <MCF>
 A:Cross-references: GB:M26199; NID:g205178; PIDN:AAAA1528.1; PID:g205179
 R:Dufau, M.L.; Minegishi, T.; Buczko, E.S.; Delgado, C.J.; Zhang, R.
 J. Steroid Biochem. 33, 715-720, 1989
 A:Title: Characterization and structure of ovarian and testicular LH/CG receptors.
 A:Reference number: A61453; MOID:90097014
 A:Accession: A61453
 A:Status: preliminary
 A:Molecule type: protein
 A:Residues: 27-33, 'X', 35-37, 'X', 39, 'X', 41-44 <DUF>
 R:Roche, P.C.; Ryan, R.J.
 J. Biol. Chem. 264, 4636-4641, 1989
 A:Title: Purification, characterization, and amino-terminal sequence of rat ovarian rece

A:Accession: A32460
 A:Molecule type: protein
 A:Residues: 27-32, 'X', 35-37 <ROC>
 C:Genetics:
 A:Insertions: 58/2, 82/2, 107/2, 132/2, 157/2, 183/2, 206/2, 231/2, 293/2, 320/2
 C:Superfamily: glycoprotein hormone receptor; leucine-rich alpha-2-glycoprotein repea
 C:Keywords: alternative splicing; G protein-coupled receptor; glycoprotein; hormone r
 Db 309 ESTVRRADNETLYSAIFEENELSGWDYDGC--SPKTLQCAPPPDAFNPCEDDMGYAF 366
 QY 374 RVEVWVAVATCGNIFVICMRPYINSEKLYAMSIISLCCADCLMGIFYVIGFPLKF 433
 F:54-77/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR1>
 Db 367 RVLIMINILAIIGNLVLFVLLTSRYKLVPRLMCLNSFPADCMGLYLLILASVDSQT 426
 QY 434 RGEYKHAQIAMESTHCOLVGSALISTEYVLLFTLEKYCIYPRPCVPGKCR- 492
 F:103-127/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR3>
 F:128-152/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR4>
 F:153-177/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR5>
 F:180-201/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR6>
 Db 427 KGOYNNHAIWMQSGGCGAAGFFVFASELVTLVTLTLEKMTITTYAVOLDKLRLH 486
 QY 493 TITVLLIMITGFIATFIPLSNKEFFKNYGTGVCPLSEDESTIGAOIYSAIFLGI 552
 F:202-226/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR7>
 F:268-389/Domain: transmembrane #status predicted <TM1>
 F:400-422/Domain: transmembrane #status predicted <TM2>
 F:444-466/Domain: transmembrane #status predicted <TM3>
 F:489-511/Domain: transmembrane #status predicted <TM4>
 F:530-551/Domain: transmembrane #status predicted <TM5>
 F:575-598/Domain: transmembrane #status predicted <TM6>
 F:610-631/Domain: transmembrane #status predicted <TM7>
 F:603-178, 199, 295, 303, 317/Binding site: carbohydrate (Asn) (covalent) #status predict
 F:647, 648/Binding site: palmitate (Cys) (covalent) #status predicted
 F:681/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicte
 Db 593 AISAARVPLI-TVTNKILLVLFYVNSCANPFLYAIPTKAFORDFLLLSRF 645
 Query Match 12.3% Score 470.5; DB 2; Length 700;
 Best Local Similarity 22.8%; Pred. No. 7,2e-24;
 Matches 149; Conservative 135; Mismatches 233; Indels 137; Gaps 23;
 QY 87 PSVSSNTVANSLOMNLRLPPDCKFNHYDLQKLDL-QNNKITSISIAFERGLNSLTKLY 145
 F:153-177/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR5>
 Db 49 PGRAGIARLSLTLVLPKVIQSAFRLNEMVYKIEISQDSLERIEANARDNLINSELL 108
 QY 146 LSNHR-ITFLKPGVFDLHLEMLIEDN-----HLRSIPFFYVNSLILLYLMNNV 198
 F:153-177/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR5>
 Db 109 IQNKNLTYTEPGAFNLPRLKYLSTGNTGIRLPDVTAKISSSEF-----NFLTEICNLH 164
 QY 199 LTRLPDRLCOHMPRLMLDLEGNHINLNLNLFISCSNLTFLVMRKKNKINHLENTEFAP 258
 F:153-177/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR5>
 Db 165 ITTIPGNAFGMNNESTLTLXGPFEEVSHAF-----NCTTILSL-----LKENLY-- 213
 QY 259 LQKIDE-----LDGSKNIENLPPLIFKDKELSQLNSLPYKIQANQFDYL 307
 F:153-177/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR5>
 Db 214 LKMHSAFOGAGPSLIDISSTYKQALPS-----HGLSIOF-----L 252
 QY 308 VKTKSLSGEIEISNIOQNFRLM-----NLSHYEFKQF--- 343
 F:153-177/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR5>
 Db 253 IALSYSLKTLTP-----SKEKFTSLVATLTTPSHCAFRMLPKKEQNFSPSIFENFSKOC 308
 A:Accession: A32460
 A:Molecule type: protein
 A:Residues: 27-32, 'X', 35-37 <ROC>
 C:Genetics:
 A:Insertions: 58/2, 82/2, 107/2, 132/2, 157/2, 183/2, 206/2, 231/2, 293/2, 320/2
 C:Superfamily: glycoprotein hormone receptor; leucine-rich alpha-2-glycoprotein repea
 C:Keywords: alternative splicing; G protein-coupled receptor; glycoprotein; hormone r
 Db 309 ESTVRRADNETLYSAIFEENELSGWDYDGC--SPKTLQCAPPPDAFNPCEDDMGYAF 366
 QY 374 RVEVWVAVATCGNIFVICMRPYINSEKLYAMSIISLCCADCLMGIFYVIGFPLKF 433
 F:54-77/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR1>
 Db 367 RVLIMINILAIIGNLVLFVLLTSRYKLVPRLMCLNSFPADCMGLYLLILASVDSQT 426
 QY 434 RGEYKHAQIAMESTHCOLVGSALISTEYVLLFTLEKYCIYPRPCVPGKCR- 492
 F:103-127/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR3>
 F:128-152/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR4>
 F:153-177/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR5>
 F:180-201/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR6>
 Db 427 KGOYNNHAIWMQSGGCGAAGFFVFASELVTLVTLTLEKMTITTYAVOLDKLRLH 486
 QY 493 TITVLLIMITGFIATFIPLSNKEFFKNYGTGVCPLSEDESTIGAOIYSAIFLGI 552
 F:202-226/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR7>
 F:268-389/Domain: transmembrane #status predicted <TM1>
 F:400-422/Domain: transmembrane #status predicted <TM2>
 F:444-466/Domain: transmembrane #status predicted <TM3>
 F:489-511/Domain: transmembrane #status predicted <TM4>
 F:530-551/Domain: transmembrane #status predicted <TM5>
 F:575-598/Domain: transmembrane #status predicted <TM6>
 F:610-631/Domain: transmembrane #status predicted <TM7>
 F:603-178, 199, 295, 303, 317/Binding site: carbohydrate (Asn) (covalent) #status predict
 F:647, 648/Binding site: palmitate (Cys) (covalent) #status predicted
 F:681/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicte
 Db 593 AISAARVPLI-TVTNKILLVLFYVNSCANPFLYAIPTKAFORDFLLLSRF 645

Thu Sep 5 14:39:29 2002

us-09-647-067-8.rpr

Page 11

Search completed: September 5, 2002, 09:01:18
Job time: 2708 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2002, 08:25:21 ; Search time 56.19 Seconds
(without alignments)
497.518 Million cell updates/sec

Title: US-09-647-067-8

Perfect score: 3834
Sequence: 1 MTSGSVFFYIIIFGKFSHG.....FTYPCMSLISQSTRLSYS 722

Scoring table: BLOSUM62
Gapop 10.0 , Gapect 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	832	21.7	1115	1	GPCR_LYMST
2	505	13.2	693	1	FSHR_CHICK
3	501	13.1	695	1	FSHR_SHEEP
4	498	13.0	695	1	FSHR_BOVIN
5	488	12.7	695	1	FSHR_PIG
6	487.5	12.7	676	1	FSHR_CALJA
7	484.5	12.6	694	1	FSHR_HORSE
8	481.5	12.6	692	1	FSHR_RAT
9	481.5	12.6	696	1	FSHR_PIG
10	481	12.5	695	1	FSHR_MACFA
11	478	12.5	695	1	FSHR_HUMAN
12	475.5	12.4	692	1	FSHR_MOUSE
13	474.5	12.4	687	1	FSHR_EQUAS
14	471.5	12.3	700	1	FSHR_MOUSE
15	470.5	12.3	700	1	FSHR_RAT
16	469	12.2	699	1	FSHR_HUMAN
17	463.5	12.1	701	1	FSHR_BOVIN
18	451.5	11.8	925	1	GLHR_ANTI
19	450	11.7	763	1	FSHR_BOVIN
20	447.5	11.7	764	1	FSHR_MOUSE
21	440.5	11.5	764	1	FSHR_SHEEP
22	437	11.4	764	1	FSHR_RAT
23	431	11.2	764	1	FSHR_HUMAN
24	430	11.2	764	1	FSHR_CANFA
25	400	10.4	538	1	FSHR_SHEEP
26	351	9.2	366	1	FSHR_MOUSE
27	302	7.9	603	1	ALS_RAT
28	296	7.7	603	1	ALS_HUMAN
29	294	7.7	603	1	ALS_MOUSE
30	293.5	7.5	905	1	TLR3_MOUSE
31	286.5	7.5	905	1	SLIT_DROME
32	286.5	7.5	1480	1	CBP8_HUMAN
33	275.5	7.2	536	1	homo sapien

34	268	7.0	361	1	CHAD_BOVIN	Q27912	bos taurus
35	260	6.8	859	1	TLR5_MOUSE	Q9J1F7	mus musculus
36	258	6.7	713	1	GAGI_HUMAN	Q753Z5	homo sapien
37	253.5	6.6	1315	1	CHAO_DROME	P12074	drosophila
38	249	6.5	331	1	PLIB_AGRBL	Q93233	agkistrodon
39	244.5	6.4	343	1	LWM_CHICK	P51890	gallus gall
40	243.5	6.4	904	1	TLR3_HUMAN	Q15455	homo sapien
41	241	6.3	567	1	GPV_MOUSE	Q08742	mus musculus
42	240	6.3	382	1	PRLP_HUMAN	P51888	homo sapien
43	235	6.1	682	1	CONN_HUMAN	Q01819	drosophila
44	235	6.1	782	1	CHAO_TRICA	P82963	tribolium c
45	233	6.1	966	1	Y918_HUMAN	Q94991	homo sapien

ALIGNMENTS

RESULT	1	STANDARD	PRT	1115 AA.
GPCR_LYMST				
ID	GPCR_LYMST			
AC	P46023			
DT	01-NOV-1995 (Rel. 32, Created)			
DT	01-NOV-1995 (Rel. 32, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	G-protein coupled receptor GRL101 precursor.			
OS	Lymnaea stagnalis (Great pond snail).			
OC	Eukaryota; Metazoa; Mollusca; Gastropoda; Pulmonata; Basommatophora;			
OC	Lymnaeidae; Lymnaea.			
OX	NCBI_TaxID=6523;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=CNS;			
RX	MEDLINE=94255418; PubMed=8197140;			
RA	Tensen C.P., van Kesteren E.R., Planta R.J., Cox K.J.A., Burke J.F.,			
RA	van Heerikhuizen H., Vreugdenhil E.;			
RT	"A G protein-coupled receptor with low density lipoprotein-binding			
RT	motifs suggests a role for lipoproteins in G-linked signal			
RT	transduction."			
RL	Proc. Natl. Acad. Sci. U.S.A. 91:4816-4820(1994)			
CC	-1- FUNCTION: MIGHT DIRECTLY TRANSDUCE SIGNALS CARRIED BY LARGE			
CC	EXTRACELLULAR (Lipo)PROTEIN(COMPLEX)'S INTO NEURONAL EVENTS.			
CC	-1- SUBCELLULAR LOCATION: Integral membrane protein.			
CC	-1- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN A SMALL NUMBER OF			
CC	NEURONS WITHIN THE CENTRAL NERVOUS SYSTEM AND TO A LESSER EXTENT			
CC	IN THE HEART.			
CC	-1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.			
CC	-1- SIMILARITY: CONTAINS 12 LDL-RECEPTOR CLASS A DOMAINS.			
CC	-1- SIMILARITY: CONTAINS 7 LEUCINE-RICH REPEATS (LRR).			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; Z23104; CAAB0651.1; -			
DR	PIR; S40241; S40241.			
DR	HSSP; Q07954; ICR8.			
DR	GCRRdb; GCR_0860; -			
DR	InterPro: IPR000276; GPCR_Rhodopsin.			
DR	InterPro: IPR002172; LDL_recept_A.			
DR	InterPro: IPR001611; LRR.			
DR	InterPro: IPR000372; LRR_Nterm.			
DR	InterPro: IPR0003592; LRR_out.			
DR	InterPro: IPR003591; LRR_tyr.			
DR	Pfam; PF00001; 7tm_1; 1.			
DR	Pfam; PF00057; LDL_recept_a; 11.			
DR	Pfam; PF00560; LRR; 6.			
DR	Pfam; PF01462; LRRNT; 1.			
DR	PRINTS; PR00237; GPCRHHODOPSIN.			
DR	PRINTS; PR00261; LDLRECEPTOR.			

DR SMART; SM00192; LDla; 12.
DR SMART; SM00370; LRR; 2.
DR SMART; SM00013; LRRNT; 1.
DR SMART; SM00369; LRR_TYP; 1.
DR PROSITE; PS00237; G_PROTEIN_RECIP_FL_1; FALSE_NEG.
DR PROSITE; PS00262; G_PROTEIN_RECIP_FL_2; 1.
DR PROSITE; PS01209; LDla_1; 6.
DR PROSITE; PS50068; LDla_2; 11.
KW G-protein coupled receptor; Transmembrane; Glycoprotein; Repeat;
leucine-rich repeat; Signal.
FT SIGNAL 1 24
FT CHAIN 25 115
FT TRANSMEM 25 767
FT TRANSMEM 768 788
FT TRANSMEM 789 801
FT TRANSMEM 802 822
FT TRANSMEM 823 857
FT TRANSMEM 858 878
FT TRANSMEM 879 887
FT TRANSMEM 888 908
FT TRANSMEM 909 941
FT TRANSMEM 942 962
FT TRANSMEM 963 988
FT TRANSMEM 989 1009
FT TRANSMEM 1010 1017
FT TRANSMEM 1018 1038
FT TRANSMEM 1039 1115
FT DOMAIN 32 523
FT DOMAIN 36 79
FT DOMAIN 77 115
FT DOMAIN 116 135
FT DOMAIN 156 196
FT DOMAIN 195 232
FT DOMAIN 231 269
FT DOMAIN 272 318
FT DOMAIN 320 363
FT DOMAIN 365 403
FT DOMAIN 404 442
FT DOMAIN 444 485
FT DOMAIN 486 525
FT REPEAT 557 579
FT REPEAT 582 605
FT REPEAT 607 629
FT REPEAT 630 653
FT REPEAT 655 677
FT REPEAT 678 701
FT REPEAT 703 725
FT DISULFID 38 53
FT DISULFID 46 66
FT DISULFID 60 77
FT DISULFID 79 91
FT DISULFID 86 104
FT DISULFID 98 113
FT DISULFID 118 131
FT DISULFID 138 153
FT DISULFID 158 170
FT DISULFID 165 183
FT DISULFID 177 194
FT DISULFID 202 220
FT DISULFID 214 230
FT DISULFID 233 245
FT DISULFID 240 256
FT DISULFID 252 267
FT DISULFID 274 291
FT DISULFID 282 304
FT DISULFID 298 316
FT DISULFID 322 339
FT DISULFID 334 352
FT DISULFID 346 361
FT DISULFID 367 379
FT DISULFID 374 392
FT DISULFID 386 401
POTENTIAL.
G-PROTEIN COUPLED RECEPTOR GRLL01.
EXTRACELLULAR (POTENTIAL).
1 (POTENTIAL).
CYTOPLASMIC (POTENTIAL).
2 (POTENTIAL).
EXTRACELLULAR (POTENTIAL).
3 (POTENTIAL).
CYTOPLASMIC (POTENTIAL).
4 (POTENTIAL).
EXTRACELLULAR (POTENTIAL).
5 (POTENTIAL).
CYTOPLASMIC (POTENTIAL).
6 (POTENTIAL).
EXTRACELLULAR (POTENTIAL).
7 (POTENTIAL).
CYTOPLASMIC (POTENTIAL).
12 x 40 AA APPROXIMATE TANDEM REPEATS
SIMILAR TO THE LDL-RECEPTOR CLASS A.
LDL-RECEPTOR CLASS A 1.
LDL-RECEPTOR CLASS A 2.
LDL-RECEPTOR CLASS A 3.
LDL-RECEPTOR CLASS A 4.
LDL-RECEPTOR CLASS A 5.
LDL-RECEPTOR CLASS A 6.
LDL-RECEPTOR CLASS A 7.
LDL-RECEPTOR CLASS A 8.
LDL-RECEPTOR CLASS A 9.
LDL-RECEPTOR CLASS A 10.
LDL-RECEPTOR CLASS A 11.
LDL-RECEPTOR CLASS A 12.
LRR 1.
LRR 2.
LRR 3.
LRR 4.
LRR 5.
LRR 6.
LRR 7.

FT DISULFID 406 418 BY SIMILARITY.
FT DISULFID 413 431 BY SIMILARITY.
FT DISULFID 425 440 BY SIMILARITY.
FT DISULFID 446 458 BY SIMILARITY.
FT DISULFID 453 474 BY SIMILARITY.
FT DISULFID 465 483 BY SIMILARITY.
FT DISULFID 488 500 BY SIMILARITY.
FT DISULFID 495 513 BY SIMILARITY.
FT DISULFID 507 523 BY SIMILARITY.
FT CARBOHYD 87 87 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 166 166 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 269 269 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 318 318 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 482 482 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 502 502 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 571 571 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 618 618 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 624 624 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 685 685 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 1115 AA; 125865 MW; E9BB01297ECE356C CRC64;
Query Match 21.7%; Score 832; DB 1; Length 1115;
Best Local Similarity 30.0%; Pred. No. 4; 1e-43;
Matches 205; Conservative 114; Mismatches 240; Indels 124; Gaps 13;
QY 27 CSLGYPCGNTKCPDLPLHNGVDDC-GNAQEDNCVY-----VLCQC-----MSLPL 75
DB 488 CLEGQFCRCR-SPCINOTKXCDGTGCLGQMDENNRVCYCPHQAICQEGVTMOCTG 546
QY 76 ELDMKRFVSVPSSVSNVATMSLQWNLIRKLPPDCFKNYHD-LQKLDLONNKITSISYA 134
DB 547 KLEKM-----PVQOEMEDLSKLMIGDNLNTSTFSATYDYKYATYDLSRHHLEPIYS 602
QY 135 FRGLNSTKLYLSHNRTEFLKPGVEDLHLEMLIEDNLSPPTFGLSLILVL 134
DB 603 FQNMWKLTHLNLADNNTSLKNG----- 625
QY 195 MNVWFLRPLDKPCQHMPLHMLDEGNHINLNLFTISCSNLTVLVYMKRKNINLHEN 254
DB 626 -----SLGLSNLQKLIHNGKNIETED 649
QY 255 TPAPLQKLELDLSGSKNIENLPLIFRKDELSQNLSTYPIOKIOANOPDYLVKLSLS 314
DB 650 TFSMHLTVLDSLNOGLTHVYKMFGLKQITVLNISRQINSINDGAFNNLANVRLID 709
QY 315 LBGLEISNIOQRMPRLMNSHIYFKKFOCYGAPHYRSCRPMTDGLSSLEMLASTIOR 374
DB 710 LSGNVTKIDQKVKFMGLPRLVELKTDTSYRCCLAPEGVKCSPKODEFSCEDLSMNVLR 769
QY 375 VFVWVVSAYVCFGNIFVYICMRPYIRSEKRLYAMSIISLCCADCLMGYLVFVIGGFDLKFR 434
DB 770 VSIWVLGVALVNGVFVIFRVKVDPRFGKHYHFLITNLALGDLMLGYLLIATADTYR 829
QY 435 GEYNKAQLMESTHCOQVSLAILSTEVSVLLTFLEKYICYIYFFRCVBPGRCT 494
DB 830 GYVISHDENMKQGLCOQAFGVSTFSELSVLTLSTLDRLCILFPLRTRGLRQAI 889
QY 495 TVLLIWTITFYAFLIPLSKFEKNTYGTNGCYCPHLSHSDTSGIAQIYSVAIFGINL 554
DB 890 IVMSCIVWLFELAVPLGLFSEYENFGSGVCLAHVPRDPRPWE-YSVGVFLILNL 948
QY 555 AAFITIVESGSMFYSV---HOSAITATELRNOVKKEMILAKRFFFIYVFDALCWPIPV 611
DB 949 LSFVLIASSTLMNF-SYAKTRSAVRAESKN-----DNAMARKMTLIWMTDECCWPIIV 1003
QY 612 VKFLSLQVEIPGTTISWVVFILPINSALNPLIYLTTRPF-----KEMIR 659
DB 1004 LGFVSLAGARADQYVAMLAFLVPLPLNSATNPVITYLTSLAPFLGNVRKRANRKRSEFIHS 1063
QY 660 F-----WYNYRORKS 669
DB 1064 FTGDTFRHSYVDGTHSYCEKKS 1086

ID	FSHR_CHICK	STANDARD:	PRT:	693 AA.
AC	P79763; Q90719;			
DT	01-NOV-1997 (Rel. 35, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Follicle stimulating hormone receptor precursor (FSH-R) (Follicotropin receptor).			
GN	FSHR.			
OS	Gallus gallus (Chicken).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;			
OC	Gallus.			
OX	NCBI_TaxID=9031;			
RP	[1]			
RP	SEQUENCE FROM N.A.			
RC	T1SSDE-Ovary;			
RC	MEDLINE=97473503; PubMed=9332357;			
RA	Marabayashi N., Suzuki A., Hoshino H., Nishimori K., Mizuno S.;			
RT	"mRNA cloning and transient expression of a chicken gene encoding a follicle-stimulating hormone receptor.";			
RT	Gene 197;1:121-127(1997).			
RL	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-WHITE LEGHORN;			
RC	MEDLINE=97057887; PubMed=8902217;			
RA	Yeo S., Bridgham J.T., Foster D.N., Johnson A.L.;			
RT	"Characterization of the chicken follicle-stimulating hormone receptor (cFSH-R) complementary deoxyribonucleic acid, and expression of cFSH-R messenger ribonucleic acid in the ovary.";			
RT	Biol. Reprod. 55:1055-1062(1996).			
RL	-1- FUNCTION: RECEPTOR FOR FOLLICLE STIMULATING HORMONE. THE ACTIVITY OF THIS RECEPTOR IS MEDIATED BY G PROTEINS WHICH ACTIVATE ADENYLATE CYCLASE.			
CC	-1- SUBCELLULAR LOCATION: Integral membrane protein.			
CC	-1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.			
CC	-1- FSH/LSH/TSH SUBFAMILY.			
CC	-1- SIMILARITY: CONTAINS 7 LEUCINE-RICH REPEATS (LRR).			
CC	-----			
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; D87871; BAA13487.1; -;			
DR	EMBL; U51097; AAC60030.1; -;			
DR	HSSP; P23945; 1XUN.			
DR	GCRDb; GCR_1077; -;			
DR	GCRDb; GCR_1137; -;			
DR	InterPro: IPR000276; GPCR_Rhodpsn.			
DR	InterPro: IPR001611; LRR.			
DR	InterPro: IPR003721; LRR_Nterm.			
DR	Pfam; PF00001; 7tm_1; 1.			
DR	Pfam; PF00560; LRR; 3.			
DR	PRINTS; PR00237; GPCRHRDOPSN.			
DR	SMART; SM00013; LRRNT; 1.			
DR	PROSITE; PS00237; G_PROTEIN_RECPT_FL_1; 1.			
DR	PROSITE; PS50262; G_PROTEIN_RECPT_FL_2; 1.			
KW	G-protein coupled receptor; transmembrane; Glycoprotein; Signal; Phosphorylation; Repeat; Leucine-rich repeat.			
FT	CHAIN	1	17	POTENTIAL.
FT	CHAIN	18	693	FOLLICLE STIMULATING HORMONE RECEPTOR.
FT	DOMAIN	18	366	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	367	387	1 (POTENTIAL)
FT	DOMAIN	388	398	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	399	421	2 (POTENTIAL)
FT	DOMAIN	422	443	EXTRACELLULAR (POTENTIAL).

FT	TRANSMEM	444	465	3 (POTENTIAL).
FT	DOMAIN	466	485	3 CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	486	508	4 (POTENTIAL).
FT	DOMAIN	509	528	4 EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	529	550	5 (POTENTIAL).
FT	DOMAIN	551	573	5 CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	574	597	6 (POTENTIAL).
FT	DOMAIN	598	608	6 EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	609	630	7 (POTENTIAL).
FT	DOMAIN	631	693	7 CYTOPLASMIC (POTENTIAL).
FT	REPEAT	45	68	LRR 1.
FT	REPEAT	69	93	LRR 2.
FT	REPEAT	95	118	LRR 3.
FT	REPEAT	119	143	LRR 4.
FT	REPEAT	168	192	LRR 5.
FT	REPEAT	193	216	LRR 6.
FT	REPEAT	218	240	LRR 7.
FT	DISULFID	442	517	BY SIMILARITY.
FT	CARBOHYD	47	47	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	191	191	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	199	199	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CONFLICT	268	268	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CONFLICT	4	4	G -> D (IN REF. 2).
FT	CONFLICT	88	88	G -> A (IN REF. 2).
FT	CONFLICT	140	140.	K -> R (IN REF. 2).
FT	CONFLICT	174	174	I -> T (IN REF. 2).
FT	CONFLICT	191	191	N -> S (IN REF. 2).
FT	CONFLICT	329	329	V -> L (IN REF. 2).
SQ	SEQUENCE	693 AA;	78697 MW;	46F9869635A1BEC CRC64;

Query Match	13.2%	Score 505;	DB 1;	Length 693;
Best Local Similarity	24.7%	Pred. No. 1.4e-23;		
Matches 1/4; Conservative	122;	Mismatches 259;	Indels 150;	Gaps 27;

```
0Y      28 SIGYEPGNITKCLPOLLHNGVDDCGNAOEDNCV-----VLCOGSLBETELDMKKPF   83
        ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db      2 SLG-----LTCILLILSCSG-----CQHHTCCLEGRAFTEQELKVQLPD-----   43
        : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY      84 TSVPSSNVAMSLOMNLIRKPDPDEKNYDLOKTLDLNKTTSISIAFRGLNSLTk   143
        : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db      44 -----IPNNATELRVLTKMKIVPRKAFTGHDEKIEISOI-----               80
        : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY      144 LYLSNRRITFLPKGVFEDJHLMLMTID -NHLSRISPPTTYGNSLLLYAMNWTRL    202
        : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db      81 -----DALEITEGVNFSSLPHKHREIKAKKLKDODAQHLPRLSYLLISTWGSIFL   135
        : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY      203 PDKPLPCQMPLRH-----WLBLEGN-HIHNRNTLETSCSNLVLY-MRKXINHLMNET    255
        : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db      136 P-----VVHKVHSQVKLLDVODNIHIERTEMTFMOJSSESVILRNKKGIOEIMDHA   189
        : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY      236 FAPLOKDELDELGSN-KIENTLPRLFCD----LKESQLMLASYNP-----IQTOANO    303
        : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db      190 FNG-"TCELEMLINDSDYNYNEKEPRKEVFOGAIGPVVIDISRTSIFLPSHGLEFIKLARS   248
        : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY      304 FDYLKL-----KSLSBEIEIS-----NQQRNERFLMNLSHYEFKKPYCG    346
        : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db      249 TYKKLKLDPVNKRFSL-TEANFYTPSHCCAFTNKRTONTMEFYICSPAKODLGDTG    306
        : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY      347 YAPHVS-----CRPNFDGISLENLSAIOR    374
        : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db      307 KRKHRSAAEDYISHYGRFGVENEFDYGLCNEVDVCSPKDFAEPCEDINGMYNLR     366
        : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY      375 VFVWVVASATPGCNLFVLCMRPIYRSEKKLYAMSI--SLCACACMGITYLFEVGEDLk   433
        : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db      367 VLIMEINITLATGNTTVLI-LISSQYKLVPPRFMCOMAFADJCIGITYLFAVSADIQ   424
        : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY      433 FGEGENKHAOLMESTHOVLGSAILESTEVSYLLFFLTLEYKICYVPRCPRPCKR   497
        : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db      425 TKSRYNNALDMOGAGCMAAGFFVFVYSLSVTYLLITERHHITIYAMQMLRWKYRLr   488
        : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY      493 -TTIVLLIWITGTIVAPIPLSKNEEFFKNYYGTWGCFPLHSEDTESTIGAOTYSVALFg   551
        : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
```

DB 485 HAVIIMVGMFAETVALLPI---FCISSMYKVSICLPHMIETPFS--QAY-VIFLAV 536
 QY 552 INIAAFIIFYSGSMYSYHOSATATETIRNGYKEMIAKRPFFIVTDAICMIPFV 611
 DB 537 LVNIAFIIICICICICIFVIRNPVVIS-----NSDTKIAKRAIIILFIDFLCMAPISF 590
 QY 612 VKFLSLQVEIPGTTISWVIFIL--PINSALNIIYTLTPRPK 654
 DB 591 FAISASLRVPLI-TVSKSKILLVLFYVINSKANCFIATFTKPR 634
 RESULT 3
 FSHR_SHEEP STANDARD; PRT; 695 AA.
 ID FSHR_SHEEP 028573; 028574; 09TS19;
 AC P35379; 028573; 028574; 09TS19; Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Follicle stimulating hormone receptor precursor (FSH-R) (Follicle
 DE receptor).
 GN FSHR.
 OS Ovis aries (Sheep).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Caprinae; Ovis.
 OC NCBI_TaxID=9940;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM FSH-R1).
 RC TISSUE-Testis;
 RX MEDLINE=93351750; PubMed=8394255;
 RA Yarney T.A., Sairam M.R., Khan H., Ravindranath N., Payne S.,
 RT Seidah N.G.;
 RT "Molecular cloning and expression of the ovine testicular follicle
 RT stimulating hormone receptor";
 RT Mol. Cell. Endocrinol. 93:219-226(1993).
 RL [2]
 RN SEQUENCE FROM N.A. (ISOFORMS FSH-R4 AND FSH-R3).
 RP STRAIN=DORSET-LEICESTER-SUPFOLK 1; TISSUE=Testis;
 RC MEDLINE=93176195; PubMed=8439338;
 RX Khan H., Yarney T.A., Sairam M.R.;
 RA "Cloning of alternatively spliced mRNA transcripts coding for variants
 RT of ovine testicular follicle receptor lacking the G protein
 RT coupling domains";
 RT Biochem. Biophys. Res. Commun. 190:888-894(1993).
 RL [3]
 RN SEQUENCE FROM N.A. (ISOFORM FSH-R2), AND CHARACTERIZATION.
 RP STRAIN=DORSET-LEICESTER-SUPFOLK 1; TISSUE=Testis;
 RC MEDLINE=98031015; PubMed=9364440;
 RX Yarney T.A., Jiang L., Khan H., MacDonald E.A., Laird D.W.,
 RA Sairam M.R.;
 RA "Molecular cloning, structure, and expression of a testicular
 RT follicle receptor with selective alteration in the carboxy terminus
 RT that affects signaling function";
 RT Mol. Reprod. Dev. 48:458-470(1997).
 RL [4]
 RN SEQUENCE FROM N.A. (ISOFORM FSH-R3), AND CHARACTERIZATION.
 RP STRAIN=DORSET-LEICESTER-SUPFOLK 1; TISSUE=Ovary;
 RC MEDLINE=20391225; PubMed=10527886;
 RX Babu P.S., Jiang L., Sairam A.M., Touyz R.M., Sairam M.R.;
 RA "Structural features and expression of an alternatively spliced growth
 RT factor type I receptor for follicle stimulating hormone in the developing
 RT ovary";
 RT Mol. Cell Biol. Res. Commun. 2:21-27(1999).
 RL [5]
 RN SEQUENCE OF 1-51 FROM N.A.
 RP MEDLINE=98031017; PubMed=9364442;
 RX Sairam M.R., Subbarayan V.S.R.;
 RA "Characterization of the 5' flanking region and potential control
 RT elements of the ovine follicle receptor gene";
 RL Mol. Reprod. Dev. 48:480-487(1997).
 CC -1- FUNCTION: Receptor for follicle stimulating hormone. The activity
 CC of isoform FSH-R1 is mediated by G proteins which activate
 CC adenylyl cyclase. Isoforms FSH-R2 and FSH-R3 also bind FSH, but

CC this does not result in activation of adenylyl cyclase. Isoform
 CC FSH-R3 may be involved in calcium signaling.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Plasma membrane
 CC (isoforms FSH-R1 and FSH-R2); Cell surface (isoform FSH-R3).
 CC -1- ALTERNATIVE PRODUCTS: 4 isoforms; FSH-R1 (shown here), FSH-R2,
 CC FSH-R3 and FSH-R4; are produced by alternative splicing.
 CC -1- TISSUE SPECIFICITY: Isoform FSH-R3 is expressed in ovary and
 CC testis, but not in kidney.
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
 CC FSH/LSH/FSH SUBFAMILY.
 CC -1- SIMILARITY: CONTAINS 6 LEUCINE-RICH REPEATS (LRR).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@isb-sdb.ch).
 CC -----
 CC EMBL; L07302; AAA31525.1; -
 CC EMBL; L12766; AAA31523.1; -
 CC EMBL; L12767; AAA31524.1; -
 CC EMBL; L36115; AAK70667.1; -
 CC EMBL; AJ131735; CA10495.1; -
 CC EMBL; AF090438; AAC61749.1; -
 CC PIR; JCI493; JCI493.
 CC HSSP; P23945; 1XUN.
 CC GCRDB; GCR_0496; -
 CC InterPro; IPR000276; GPCR_Rhodpsn.
 CC InterPro; IPR001611; LRR.
 CC InterPro; IPR000372; LRR_Nterm.
 CC Pfam; PF00001; 7tm.1; 1.
 CC Pfam; PF00560; LRR; 5.
 CC Pfam; PF01462; LRRNT; 1.
 CC PRINTS; PR00373; GLYCHROMER.
 CC PRINTS; PR01143; FSHRECEPTOR.
 CC SMART; SM00013; LRRNT; 1.
 CC PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
 CC PROSITE; PS00262; G_PROTEIN_RECEP_F1_2; 1.
 CC G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;
 CC Phosphorylation; Repeat; Leucine-rich repeat; Alternative splicing.
 CC KW CHAIN 18 695
 CC FT SIGNAL 1 17
 CC FT DOMAIN 18 366
 CC FT TRANSMEM 367 387
 CC FT DOMAIN 388 398
 CC FT TRANSMEM 399 421
 CC FT DOMAIN 422 443
 CC FT TRANSMEM 444 465
 CC FT DOMAIN 466 485
 CC FT TRANSMEM 486 508
 CC FT DOMAIN 509 528
 CC FT TRANSMEM 529 550
 CC FT DOMAIN 551 573
 CC FT TRANSMEM 574 597
 CC FT DOMAIN 598 608
 CC FT TRANSMEM 609 630
 CC FT DOMAIN 631 695
 CC FT REPEAT 44 68
 CC FT REPEAT 69 93
 CC FT REPEAT 119 192
 CC FT REPEAT 170 193
 CC FT REPEAT 193 216
 CC FT REPEAT 218 240
 CC FT DISULFID 442 517
 CC FT CARBOHYD 191 191
 CC FT CARBOHYD 199 199
 CC FT CARBOHYD 293 293
 CC FT VARSLIC 126 133
 CC FT VARSLIC 135 695
 CC FT VARSLIC 224 259
 CC VARSLIC 135 695
 CC VARSLIC 224 259
 CC -----
 CC POTENTIAL.
 CC FOLLICLE STIMULATING HORMONE RECEPTOR.
 CC EXTRACELLULAR (POTENTIAL).
 CC 1 (POTENTIAL).
 CC CYTOPLASMIC (POTENTIAL).
 CC 2 (POTENTIAL).
 CC EXTRACELLULAR (POTENTIAL).
 CC 3 (POTENTIAL).
 CC CYTOPLASMIC (POTENTIAL).
 CC 4 (POTENTIAL).
 CC EXTRACELLULAR (POTENTIAL).
 CC 5 (POTENTIAL).
 CC CYTOPLASMIC (POTENTIAL).
 CC 6 (POTENTIAL).
 CC EXTRACELLULAR (POTENTIAL).
 CC 7 (POTENTIAL).
 CC CYTOPLASMIC (POTENTIAL).
 CC LRR 1.
 CC LRR 2.
 CC LRR 3.
 CC LRR 4.
 CC LRR 5.
 CC LRR 6.
 CC BY SIMILARITY.
 CC N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC L1SNGIK -> EKRRNRNI (IN ISOFORM
 CC FSH-R4).
 CC MISSING (IN ISOFORM FSH-R4).
 CC DISRRIRSLPSYGEENIKKRAKSYHLKLPISLE ->


```

FT      SPLHCHAHLOSEFFVYVCGOREHISEFGKLSKQHPN (IN
FT      ISOFORM FSH-R3).
FT      VARSPLIC 260 695
FT      VARSPLIC 643 670
FT      VARSPLIC 671 695
FT      VARSPLIC 695 AA; 78237 MW; BFB75D9D8C0D4B CRC64;
SO      SEQUENCE

Query Match      13.1%; Score 501; DB 1; Length 695;
Best Local Similarity 24.4%; Pred. No. 2.5e-23;
Matches 168; Conservative 130; Mismatches 245; Indels 146; Gaps 28;

QY      40 CLPDLHC-NGVDCCGAGDEDCNVVLCOCMSLPGLELDMKRFYSPS-VSSNVYAMS 97
      18 CHHRLCHCSNG-----VFLLCQDSKY-----TEPSPDLPRDAVELR 52
QY      98 LQWMLRLKPPDCFRKNYDLOKLDLQNNKITSISYAFRGINSILTKYLSHNRITFLKPG 157
      53 FVLKRLVIRPEGAFSGFGDLEKIEISQNDVLEY-----LEAN 89
QY      158 VFEDLHRLKLEMLIED-NHLSRISPPRTYGLNSLLVLMANNVLRLPDKPLCQHPRLHW 216
      90 VFSMLPLKHEIRIKANNLLYIDDAFQNLPLERYLLISNTGIRHLPVAKHQISLOKV-L 148
QY      217 LDLEGN-HIHNLRMLTFISCSNLFVLY-MRKNKINHLENTFAPLOKDELGLSN-KIE 273
      149 LDIDNNINIHVERNSPFGSLFESMIYWLKNGIOETHNCAFNGTQ-LDELINLSDNSMLE 207
QY      274 NLPLPLFRD-----LKELSQNLISYNP-----IQKIQANQFDLYLKLSLS-LEGIEI 320
      208 ELPNDVFGAGSAPYLLDISRTIRSLPSYGLNLEKILAKSTYHLKPLSEKFPVTLVEA 267
QY      321 S-----NIQORMFRPLMNS-----H 336
      268 SLTYPSHCCAFAMNRQTSIDLPHICNKSILRQEVDMTQARGORISLAEDDEPSYAKGFD 327
QY      337 IYFKKFOY-CGYAPHYKSCKPNTDGISLENLASIIIOYFVWVWSAVTCFNGNIEVICM 394
      328 MMYSEFVDLCSEYVDV-TCSPEPDAPPCEDIMGYDLRLVINFISLATGNILVYI 386
QY      395 RPYRSEKNLYAMSI--SLCCADCLMGIVLFGVGGDLKRGEXKNKHAQIMESTHCQL 452
      387 -LITSQYKLVPRFLMCNLAFAADLCIGIYLLIASVYVHKSYQHNVAIDMOCAGCDA 444
QY      453 VGSALISTEVSULLTFLTEKICIVYPRCYRPGCR-----TIVLLILMTGITV 507
      445 AGEFTVASELSVYTLTAITLERMHTTTHAMQL-----ECKYVHRAASIMLVGWFAFAV 500
QY      508 AFILSNKEFFKNYGTGVCFPPLHSEDTESIGAQIYSVAFLGINLAFTIIVSYGSM 567
      501 ALFPI-----FGISSYMKYSICLPM---DIDSPLSOLYMSL-LVINATLAFVYICGCTYHI 552
QY      568 FYSVHQATATEIRNOYKEMILAKREFITFDALCMWIDIFVYFKLSLLOVEIPGIT 627
      553 YLTVANPNITSS-----SDTKIAKRAMMLTFEDLCMAPISEFAISASIKVPLI-TVS 605
QY      628 SWVYIFIL--PINSALNPILYTLITRPK 654
      606 KSKILLVLEYPINSCANFLYAIFTRNR 634

RESULT 4
FSHR_BOVIN STANDARD; PRT; 695 AA.
AC P35376;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Follicle stimulating hormone receptor precursor (FSH-R) (Folliotropin
  receptor).
GN FSHR.
OS Bos taurus (Bovine).

```

```

CC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC      Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
CC      Bovidae; Bovinae; Bos.
CC      NCBI_TaxID=9913;
CC      [1]
CC      SEQUENCE FROM N.A.
CC      STRAIN=HOLSTEIN; TISSUE=Ovary, and Testis;
CC      MEDLINE=95127199; PubMed=782612;
CC      RA Houde A., Lambert A., Saumande J., Silversides D.W., Lussier J.G.;
CC      RT "Structure of the bovine follicle-stimulating hormone receptor
CC      complementary DNA and expression in bovine tissues.";
CC      RL Mol. Reprod. Dev. 39:127-135(1994).
CC      CC -1- FUNCTION: RECEPTOR FOR FOLLICLE STIMULATING HORMONE. THE ACTIVITY
      OF THIS RECEPTOR IS MEDIATED BY G PROTEINS WHICH ACTIVATE
      ADENYLATE CYCLASE.
      CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
      CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
      CC FSH/LSH/TSH SUBFAMILY.
      CC -1- SIMILARITY: CONTAINS 6 LEUCINE-RICH REPEATS (LRR).
      CC -----
      CC This SWISS-PROT entry is copyright. It is produced through a collaboration
      CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
      CC the European Bioinformatics Institute. There are no restrictions on its
      CC use by non-profit institutions as long as its content is in no way
      CC modified and this statement is not removed. Usage by and for commercial
      CC entities requires a license agreement (See http://www.isb.ch/announce/
      CC or send an email to license@isb.ch).
      CC -----
      CC EMBL; L22319; AAC37324.1; -.
      CC DR HSSP; P23945; 1XUN.
      CC DR GCRDB; GCR_0766; -.
      CC DR InterPro: IPRO00276; GPCR_Rhodpsn.
      CC DR InterPro: IPRO01611; LRR.
      CC DR InterPro: IPRO00372; LRR_Nterm.
      CC DR Pfam; PF00001; 7tm_1; 1.
      CC DR Pfam; PF00560; LRR; 5.
      CC DR Pfam; PF01462; LRRNT; 1.
      CC DR PRINTS; PRO0373; GLYCHROMONER.
      CC DR PRINTS; PRO1143; FSHRECEPTOR.
      CC DR SMART; SM00013; LRRNT; 1.
      CC DR PROSITE; PS00237; G_PROTEIN_RECPEP_F1_1; 1.
      CC DR PROSITE; PS0262; G_PROTEIN_RECPEP_F1_2; 1.
      CC KW G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;
      CC phosphorylation; Repeat; Leucine-rich repeat.
      CC FT SIGNAL 1 17
      CC FT CHAIN 1 695
      CC FT DOMAIN 18 366
      CC FT TRANSMEM 367 387
      CC FT DOMAIN 388 398
      CC FT TRANSMEM 399 421
      CC FT DOMAIN 422 443
      CC FT TRANSMEM 444 465
      CC FT DOMAIN 466 485
      CC FT TRANSMEM 486 508
      CC FT DOMAIN 509 528
      CC FT TRANSMEM 529 550
      CC FT DOMAIN 551 573
      CC FT TRANSMEM 574 597
      CC FT DOMAIN 598 608
      CC FT TRANSMEM 609 630
      CC FT DOMAIN 631 695
      CC FT REPEAT 69 68
      CC FT REPEAT 69 93
      CC FT REPEAT 119 143
      CC FT REPEAT 170 192
      CC FT REPEAT 193 216
      CC FT REPEAT 218 240
      CC FT DISULFID 442 517
      CC FT CARBOHYD 191 191
      CC FT CARBOHYD 199 199
      CC FT CARBOHYD 293 293
      CC SO SEQUENCE 695 AA; 78084 MW; 18F9DFEFC046380D CRC64;

```


FT CONFLICT 60 60 V -> A (IN REF. 1).
 FT CONFLICT 166 166 V -> M (IN REF. 1).
 FT CONFLICT 215 215 Q -> H (IN REF. 1).
 FT CONFLICT 247 247 K -> R (IN REF. 1).
 FT CONFLICT 257 257 S -> T (IN REF. 1).
 FT CONFLICT 334 334 D -> N (IN REF. 1).
 FT CONFLICT 349 349 E -> K (IN REF. 1).
 FT CONFLICT 352 352 T -> A (IN REF. 1).
 FT CONFLICT 383 383 V -> E (IN REF. 1).
 FT CONFLICT 407 407 A -> T (IN REF. 1).
 FT CONFLICT 421 421 V -> I (IN REF. 1).
 FT CONFLICT 427 427 T -> S (IN REF. 1).
 FT CONFLICT 435 435 D -> N (IN REF. 1).
 FT CONFLICT 483 483 L -> V (IN REF. 1).
 FT CONFLICT 550 550 T -> I (IN REF. 1).
 FT CONFLICT 586 586 A -> V (IN REF. 1).
 FT CONFLICT 607 607 S -> L (IN REF. 1).
 FT CONFLICT 691 691 R -> H (IN REF. 1).
 SQ SEQUENCE 695 AA; 78172 MM; E9EBDB29C79C450 CRC64;

Query Match 12.7%; Score 488; DB 1; Length 695;
 Best Local Similarity 23.8%; Pred. No. 1.5e-22;
 Matches 164; Conservative 127; Mismatches 253; Indels 144; Gaps 26;

OY 40 CLPOLHC-NGVDGCGNADBCNVCVVCOCMSLPGLELDMKKPPTSVPSVNTAMSL 98
 DB 18 CHRRCHCSNG-----VFICQESKYTEIP-----PDLPRAVNEIRF 53
 OY 99 QWNLIRKLPPDCKFNKNDLQKLDONNKITSISIAFRGLSLTKLYLSHNRIFLAKPGV 158
 DB 54 VLTKRVRPKGFSGFGLKEIKESQNDVLE-----IEANV 90
 OY 159 FEDHREMLIIE-D-NHLSRISPPFTYGLNSLILLYLMANNVLTPLRDPKPLCOHMRPLML 217
 DB 91 FSNLPKLEIRIEKANNNLLYIDPDAFQMLPNLRVLLISNTGKXHPAVHKIOSLOKV-L 149
 OY 218 DLEGH-HIHNLNRTFISCS-NLTVLMRKNNKINHLNENTFAPLOKLELDLGSN-KTEN 274
 DB 150 DIODINNIHTVERNSFVLSFESMTLMSKNGIREIHNCAGFTQ-LDELNLSDNDLEE 208
 OY 275 LPPLIFKDLKELISQNLNSYNPQIOANQFDVLVLAKSLS-----LRGIE----- 319
 DB 209 LPNDVFGASGVILLIDIRTRHSLPSYGLELTKLRKSTYNLKKLSLEKFTVLMES 268
 OY 320 -----ISNIQRM--FRPLNLSHI----- 337
 DB 269 LTPYSHCCAFANMRROISDLHPICNKSILRQEVDMTQARGORVSLAEDGESSLAKEPDT 328
 OY 338 YFKKQY--CGAPHYRSCKPTDGISSLENLNASIIQRFVWVWVSATVCGENIFVICMR 395
 DB 339 MYSEDYDLCKNEVDV-ICSEPDFTFNCEDMGHDLRLVMTFISLITAGNITVLVI- 386
 OY 396 PYIRSENKLYMSII--SLCCADCLMGITYLVIGFDFLKFREYKHKHOLMESTHCOLV 453
 DB 387 -LITSQYLYTVRFMLCNLAFRADLCIGIYLLILASVDHTKQYINNYALDMOTGACDDAA 445
 OY 454 GSLALISTEVSALLFTLEKYICIVYPRCVRPKCR-----TIVLILITWITGFIYA 508
 DB 446 GFFYVFASELASYVLTALTALERMHTITHAMQL-----QCKVOLRHAASIMLVGMIFAFTVA 501
 OY 509 FPLPSNKKEFFKNYGTNCVCPFLSHEDTESIGAQIYVAIFGLINLAFTIIVESYGSNF 568
 DB 502 LFPV---FGISSYMKVSIQPM--DIDSPLSQLYVSL-LVLNVLAFFVIOGQYTHY 553
 OY 569 YSVHOSATATATFIRNOVKEMILAKRFFIYFTDALCWPPIPVVFKLSLLOVEIGTITS 628
 DB 554 LTVRPNMSS-----SSDTKIARKMALITPDLCAPISFALISAKLPVL-TYSK 606
 OY 629 WVVFIL--PINSALNPLIYLTITTRPFK 654
 DB 607 SKILLVLFYPINSCANPFLYAFETKNR 634

RESULT 6
 LSHR_CALJA STANDARD; PRT: 676 AA.
 AC 002721;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Lutropin-choriogonadotropic hormone receptor precursor (LH/CG-R)
 DE (LSH-R) (lutalinizing hormone receptor).
 GN LHCR.
 OS Callithrix jacchus (Common marmoset).
 OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 CC Mammalia; Euthera; Primates; Platyrrhini; Callitrichidae;
 CC Callithrix.
 OX NCBI_TaxID=9483;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Testis;
 RX MEDLINE=97307697; PubMed=9165039;
 RA Zhang F.-P., Rannikko A.S., Manna P.R., Fraser H.M., Huhtaniemi I.T.;
 RT "Cloning and functional expression of the lutalinizing hormone
 RT receptor complementary deoxyribonucleic acid from the marmoset monkey
 RT testis: absence of sequences encoding exon 10 in other species.";
 RL Endocrinology 138:2481-2490(1997).
 CC -1- FUNCTION: RECEPTOR FOR LUTROPIN-CHORIOGONADOTROPIC HORMONE.
 CC THE ACTIVITY OF THIS RECEPTOR IS MEDIATED BY G PROTEINS WHICH
 CC ACTIVATE ADENYLATE CYCLASE.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
 CC FSH/LSH/TSH SUBFAMILY.
 CC -1- SIMILARITY: CONTAINS 5 LEUCINE-RICH REPEATS (LRR).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: U80673; AAB53698.1; -.
 DR HSSP: P22888; 1IUT.
 DR GCRdb, GCR_1360; -.
 DR InterPro: IPR000276; GPCR_Rhodpsn.
 DR InterPro: IPR001611; LRR.
 DR InterPro: IPR000372; LRR_Nterm.
 DR Pfam: PF00001; 7tm_1; 1.
 DR Pfam: PF00560; LRR; 2.
 DR PRINTS: PR00373; GLYCHROMONER.
 DR PRINTS: PR01144; LSHRECEPTOR.
 DR SMART: SM00013; LRRNT; 1.
 DR PROSITE: PS00237; G-PROTEIN_RECCEP_F1_1; 1.
 DR PROSITE: PS50262; G-PROTEIN_RECCEP_F1_2; 1.
 KW G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;
 KW Phosphorylation; Repeat; Leucine-rich repeat.
 FT SIGNAL 1 29
 FT CHAIN 30 676
 FT DOMAIN 30 340
 FT DOMAIN 341 362
 FT DOMAIN 363 372
 FT DOMAIN 373 393
 FT DOMAIN 394 416
 FT DOMAIN 417 439
 FT DOMAIN 440 459
 FT DOMAIN 460 482
 FT DOMAIN 483 502
 FT DOMAIN 503 526
 FT DOMAIN 527 547
 FT DOMAIN 548 571
 FT DOMAIN 572 582
 FT TRANSMEM 583 604
 FT TRANSMEM 604

FT DOMAIN 605 676 CYTOPLASMIC (POTENTIAL).
 FT REPEAT 52 75 LRR 1.
 FT REPEAT 126 151 LRR 2.
 FT REPEAT 153 175 LRR 3.
 FT REPEAT 176 200 LRR 4.
 FT REPEAT 225 248 LRR 5.
 FT REPEAT 416 491 BY SIMILARITY.
 FT CARBOHYD 103 103 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 178 178 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 199 199 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 676 AA; 75677 MM; FB01822764154E1 CRC64;

Query Match 12.7%; Score 487.5; DB 1; Length 676;
 Best Local Similarity 23.8%; Pred. No. 1.6e-22;
 Matches 144; Conservative 133; Mismatches 292; Indels 37; Gaps 14;

QY 59 EDNCVAVLCQMSLPGLFELDMKRPSTVSVSNVAMSLOMNLIRKLPPDCFKNYHDLQ 118
 DB 30 EARCPCPCPCCTPDGALF-----CPGCGAGLRLSLAYLPVKVIPSQAFRLNEVI 80
 QY 119 KLDF-QNNKITSISYAFRGINSILTKLYSHNR-ITFLKPGVFEDLHLEWLIIDNHLS 176
 DB 81 KIEISQSDSLERIANAFDNLINSEILIQNTKLIHIEPGATNLPRKLTSICNTGIR 140
 QY 177 RISPT--FYGLNSILLVNNVTLRLDKPLCOHMRMLHMLDEGNHINLRNLPFIS 234
 DB 141 KFPDYTKFSESTNFIIEICDNLHTTIPGNAPGCMNESITLKYGNGFEVQSHAFNG 200
 QY 235 CSNLTVLMKRNKINHLNENFAPLQKIDELDLGSKNIENPLIFKLKEL---SQNL 291
 DB 201 TTVISLVAKENVHLERHNGAFRGATGSPILDISFTKLQALPSSHGLSIOQLIATSYSL 260
 QY 292 SYNPIOKIQANOFDYLVKLSLEGISNIOQMFRLMNLSHIYKPKVOYCGAPHV 351
 DB 261 KILSREKFAHLDATLTTPS---HCCAFRVPTKDYAIPTAESQSDWDYDGHLPKT 317
 QY 352 RSCRPNDIGISLENLASSIQRFVWVVSATGEGNIFVICMREPIYSENKLYAMSIIS 411
 DB 318 PCRAPEPAFNPCEIDIMGYDLRLVILNITAIIMGNMTVEFLITSRYKLVPRFLMCN 377
 QY 412 LCCADCLMGITLVFYIGGDLKFRGEYKHAQDMESTCOLVGSALISTEVSULLFL 471
 DB 378 LSPDFCGLTLLLASVDSQTKGYNHAIQDMQSGCTNAGFTVASELSYTLTLLVI 437
 QY 472 TLEKXICVYFRCVPRPKCR-TIVLILIVITGFIYAFIPLSNKEFKNYGTNGVCFP 530
 DB 438 TLERHHTTTVAIHLDQKRLRHALIMGLWFLSLIAMLPLVG---VSNTMKYS-ICFP 493
 QY 531 LHSDETSIGAQIYSVAIFGICINLAFTIIVSYGSMFYSVHQSATVATEIRNOVKREMI 590
 DB 494 M---DVETTLQIYILITLI-LNVAFIICACYKIKIYFAVNPDELAMAT-----NKDTK 543
 QY 591 LAKREFEIVFDALCWPIPFVVKFSLQVLEPGITISWVYIFIL--PINSALNPILTL 648
 DB 544 IAKKAAIILIFDTOMAFISPAISAFKMPIL-TVINSKVLVLEYPINSCANPELAI 602
 QY 649 TTRPEK 654
 DB 603 FTKTRF 608
 RESULT 7
 FSHR HORSE STANDARD; PRT; 694 AA.
 AC P47799;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Follitole stimulating hormone receptor precursor (FSH-R) (Follitropin receptor).
 GN FSHR.
 OS Equus caballus (Horse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
 NCBI_TaxID=9796;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Testis;
 RA MEDLINE=94256980; PubMed=8198575;
 RA Robert P., Amselem S., Christophe S., Benfla J.L., Bellet D.,
 RA Koman A., Bidart J.M.;
 RT "Cloning and sequencing of the equine testicular follitropin
 RT receptor."
 RL Biochem. Biophys. Res. Commun. 201:201-207(1994).
 CC -1- FUNCTION: RECEPTOR FOR FOLLICLE STIMULATING HORMONE. THE ACTIVITY
 CC OF THIS RECEPTOR IS MEDIATED BY G PROTEINS WHICH ACTIVATE
 CC ADENYLATE CYCLASE. AMONG ALL MAMMALIAN FSH RECEPTORS, ON THE HORSE
 CC RECEPTOR DOES NOT BIND LH/CHORIONIC GONADOTROPHIN (CG).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
 CC -1- FSH/LSH/TSR SUBFAMILY.
 CC -1- SIMILARITY: CONTAINS 6 LEUCINE-RICH REPEATS (LRR).
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed, usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; S70150; AAB30854.1; -.
 DR HSSP; P23945; IXUN.
 DR GCRDB; GCR1251; -.
 DR InterPro: IPR000276; GPCR_Rhodpsn.
 DR InterPro: IPR001611; LRR.
 DR InterPro: IPR000372; LRR_Nterm.
 DR Pfam; PF00001; 7tm_1; 1.
 DR Pfam; PF00560; LRR; 4.
 DR Pfam; PF01462; LRRNT; 1.
 DR PRINTS; PR00237; GPCRHDOPSN.
 DR SMART; SM00013; LRRNT; 1.
 DR PROSITE; PS00237; G-PROTEIN RECP_F1_1; 1.
 DR PROSITE; PS50262; G-PROTEIN_RECP_F1_2; 1.
 KW G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;
 KW Phosphorylation; Repeat; Leucine-rich repeat.
 FT SIGNAL 1
 FT CHAIN 18 694
 FT DOMAIN 18 365
 FT TRANSSEM 366 386
 FT DOMAIN 387 397
 FT TRANSSEM 398 420
 FT DOMAIN 421 442
 FT TRANSSEM 443 464
 FT DOMAIN 465 484
 FT TRANSSEM 485 507
 FT DOMAIN 508 527
 FT TRANSSEM 528 549
 FT DOMAIN 550 572
 FT TRANSSEM 573 596
 FT DOMAIN 597 607
 FT TRANSSEM 608 629
 FT DOMAIN 630 694
 FT REPEAT 44 68
 FT REPEAT 69 93
 FT REPEAT 119 143
 FT REPEAT 168 192
 FT REPEAT 193 216
 FT REPEAT 218 240
 FT DISULFID 441 516
 FT CARBOHYD 191 191
 FT CARBOHYD 199 199
 FT CARBOHYD 268 268
 FT CARBOHYD 293 293
 SQ SEQUENCE 694 AA; 78004 MM; E2F077C5E8BCA54 CRC64;

FT REPEAT 119 143 LRR 3.
 FT REPEAT 170 192 LRR 4.
 FT REPEAT 193 216 LRR 5.
 FT REPEAT 218 240 LRR 6.
 FT DISULEID 442 517 BY SIMILARITY.
 FT CARBOHYD 191 191 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 199 199 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 293 293 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 318 318 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SO SEQUENCE 695 AA; 78343 MW; 0D60A233729B5250 CRC64;

Query Match 12.5%; Score 481; DB 1; Length 695;

Best Local Similarity 24.8%; Pred. No. 4e-22;

Matches 158; Conservative 125; Mismatches 238; Indels 116; Gaps 25;

QY 99 QWNLIRKLPDCCFKNYNDLQKLDQNNKNTSISYAFRGSLNLRKLYSHNRI-TFLKPG 157
 DB 33 QESKVTETIPSDLPNNAIELR--FVHTKRLRYIQKAFSGFDELKIEISQNDVLEVIAD 89
 QY 158 VFEDLHLEWLIID-NHLRSPPTFYGLNSLILVLMNVNLRPLDPCQHMPLRH- 215
 DB 90 VFSNLPRLHERIKANNLLYINPEAFONLPNLKYLISNTGIKHPD-----VHKIHS 143
 QY 216 ---WLDLEGN-HIHLNRLNLFISCS- NLTVLVNRKKINHLNENTPAPLOKDELDC- 268
 DB 144 FOKVLIDIODINIHITERNSEFVGLSEFESVILMLNKNGIOEIHNCAGFTQ-LDELNLSD 202
 QY 269 SNKLENPPLIFKRLKLSQNLNLYNPLOKQANOFDLYVTKLS-----LEGIE--- 319
 DB 203 NNNLELPNDVFHGAASPVLLIDSTRHSLPSYGLEMLKRLARSTYNNLKLPSLEKLV 262
 QY 320 -----ISNIOORM--FRPLMNS----- 335
 DB 263 ALMEASLTYPSCAFANMRQISELHPICMKTILROEVNMTQTRGRSSLAEDNESSY 322
 QY 336 ---HIYFKKFOY--CGYAPHRSCRPNTDGISSLENLASIIORVFWVVSATCFGNI 389
 DB 323 SRGDMTYAERDYDLCEVNDV--TCSPPKPAFNCECDILGYNLRILWIFISIIATGNI 381
 QY 390 FVTCMRPIYSENKLYAMSI-----ISLCACDCLMGIIYIPYIGFDLKFREYKKAOLW 444
 DB 382 IVL-----VLTTSQYKLYVRFELMCNLAFFDLGIGIYLLIASVDIHTKQSYHNAYADW 436
 QY 445 MESTHCCVGLSALITSEVSULLTFLLEKICIVYPERCVRGKCR-----PIYVLI 499
 DB 437 QTGAGCAAGFTYFASSELVYITLITALEKMHITAMOL-----DCKVHRHAASVWM 492
 QY 500 IWTGTVAEPLISNKEFFKNYGTNGVCEPLHSEDTESIGAOITYVAIFLGIMAAFI 559
 DB 493 GMIFAFMAALEPI---FGISSYMKVASICPM---DIDSPLSQLYVMSL-LVLNVLAFFV 544
 QY 560 IVFYSGMFSYHOSATATAEIRNOVKKEMILARFFIYVTDLCMIPYVAFSLIQ 619
 DB 545 ICGGYTHIYLVNPNVTSS-----SSDRIRARMMALJFTDPLCMAPISFPAISASLK 598
 QY 620 VEIPGTTSMVVFIL--PINSALPILYTLTTPPK 654
 DB 599 VPLI-TVSKAKIILVLYFPINSCANPFLYALFTNFR 634
 RESULT 11
 FSHR_HUMAN
 AC P23945; STANDARD; PRT; 695 AA.
 DT 01-MAR-1992 (Rel. 21, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Follicle stimulating hormone receptor precursor (FSH-R) (Follictropin
 receptor).
 GN FSHR.
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Ovary;
 RX MEDLINE=9122171; PubMed=1709010;
 RA Minegishi T., Nakamura K., Takakura Y., Ibuki Y., Igarashi M.;
 RT Cloning and sequencing of human FSH receptor cDNA."
 RL Biochem. Biophys. Res. Commun. 175:1125-1130(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Testis;
 RX MEDLINE=93246012; PubMed=1301382;
 RA Kelton C.A., Cheng S.Y., Nugent N.P., Schweickhardt R.L.,
 RA Rosenthal J.L., Overton S.A., Wands G.D., Kuzeja J.B., Luchette C.A.,
 RA Chappel S.C.;
 RT "The cloning of the human follicle stimulating hormone receptor and
 its expression in COS-7, CHO, and Y-1 cells."
 RL Mol. Cell. Endocrinol. 89:141-151(1992).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Tilly L.T., Alhara T., Nishimori K., Jai X.-C., Billig H.,
 RA Kowalski K.I., Perlas E.A., Hsueh A.J.;
 RL Submitted (XXX-1992) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE OF 1-342 FROM N.A.
 RC TISSUE-Testis;
 RX MEDLINE=93075197; PubMed=1359889;
 RA Gromoll J., Gudermann T., Nieschlag E.;
 RT "Molecular cloning of a truncated isoform of the human follicle
 stimulating hormone receptor."
 RL Biochem. Biophys. Res. Commun. 188:1077-1083(1992).
 RN [5]
 RP SEQUENCE OF 1-51 FROM N.A.
 RX MEDLINE=95011044; PubMed=7926278;
 RA Gromoll J., Dankbar B., Gudermann T.;
 RT "Characterization of the 5' flanking region of the human follicle-
 stimulating hormone receptor gene."
 RL Mol. Cell. Endocrinol. 102:93-102(1994).
 RN [6]
 RP 3D-STRUCTURE MODELING OF 49-228.
 RX MEDLINE=96363672; PubMed=8747461;
 RA Jiang X., Dreano M., Buckler D.R., Cheng S., Ythier A., Wu H.,
 RA Handrickson W.A., el Taylor N.;
 RT "Structural predictions for the ligand-binding region of glycoprotein
 hormone receptors and for the nature of hormone-receptor interactions."
 RL Structure 3:1341-1353(1995).
 CC -1- FUNCTION: RECEPTOR FOR FOLLICLE STIMULATING HORMONE. THE ACTIVITY
 OF THIS RECEPTOR IS MEDIATED BY G PROTEINS WHICH ACTIVATE
 CC ADENYLATE CYCLASE.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A LONG FORM (SHOWN HERE) AND A
 SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: SEROTOLI CELLS AND OVARIAN GRANULOSA CELLS.
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
 CC FSH/LSH/TSH SUBFAMILY.
 CC -1- SIMILARITY: CONTAINS 5 LEUCINE-RICH REPEATS (LRR).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M65085; AAA52477.1; -
 CC EMBL; S59900; AAB26480.1; -
 CC EMBL; M95489; AAA52478.1; -
 CC EMBL; X68044; CAA48179.1; -
 CC EMBL; S73199; AAB32071.1; -
 CC PIR; JN0122; JN0122.
 CC PDB; 1XUN; 15-MAY-97.


```

DR GCRDb: GCR_0071: -
DR GCRDb: GCR_0404: -
DR GCRDb: GCR_0588: -
DR GCRDb: GCR_0690: -
DR MIM: 136435: -
DR InterPro: IPRO00276: GPCR_Rhodopsin.
DR InterPro: IPRO01611: LRR.
DR InterPro: IPRO00372: LRR_Nterm.
DR Pfam: PF00001: 7tm_1; 1.
DR Pfam: PF00560: LRR; 4.
DR Pfam: PF01462: LRRNT; 1.
DR PRINTS: PRO0373: GSHCHROMER.
DR PRINTS: PRO1143: GSHCHROMER.
DR SMART: SM00013: LRRNT; 1.
DR PROSITE: PS00013: G-PROTEIN_RECEP_F1_1; 1.
DR PROSITE: PS0262: G-PROTEIN_RECEP_F1_2; 1.
DR G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;
KW Phosphorylation; Repeat; Leucine-rich repeat; Alternative splicing;
FT 3D-structure.
FT SIGNAL 1 17 POTENTIAL.
FT CHAIN 18 695 FOLLICLE STIMULATING HORMONE RECEPTOR.
FT DOMAIN 18 366 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 367 387 1 (POTENTIAL).
FT DOMAIN 388 398 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 399 421 2 (POTENTIAL).
FT DOMAIN 422 443 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 444 465 3 (POTENTIAL).
FT DOMAIN 466 485 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 486 508 4 (POTENTIAL).
FT DOMAIN 509 528 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 529 550 5 (POTENTIAL).
FT DOMAIN 551 573 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 574 597 6 (POTENTIAL).
FT DOMAIN 598 608 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 609 630 7 (POTENTIAL).
FT DOMAIN 631 695 CYTOPLASMIC (POTENTIAL).
FT REPEAT 69 93 LRR 1.
FT REPEAT 119 143 LRR 2.
FT REPEAT 170 192 LRR 3.
FT REPEAT 193 216 LRR 4.
FT REPEAT 218 240 LRR 5.
FT DISULFID 442 517 BY SIMILARITY.
FT CARBOHYD 191 191 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 199 199 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 293 293 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 318 318 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 224 285 MISSING (IN SHORT ISOFORM).
FT VARSPLIC 342 695 MISSING (IN SHORT ISOFORM).
FT CONFLICT 13 13 S -> R (IN REF. 4).
FT CONFLICT 112 112 N -> T (IN REF. 1).
FT CONFLICT 197 198 EL -> AV (IN REF. 1).
FT CONFLICT 295 295 S -> P (IN REF. 4).
FT CONFLICT 307 307 T -> A (IN REF. 1).
FT CONFLICT 307 680 N -> S (IN REF. 1).
SQ SEQUENCE 695 AA; 78294 MM; 723B8E71F76D2CD5 CRC64;

```

Query Match 12.5%; Score 478; DB 1; Length 695;
 Best Local Similarity 23.1%; Pred. No. 6e-22;
 Matches 160; Conservative 129; Mismatches 250; Indels 154; Gaps 26;

```

OY 40 CLPOLHNGVDDGQNDENGVVLLCOGMSLPLELDMKKPFTSVS-VSSNTAMSL 98
DB 18 CHHRCHCSN-----RFLDQ-----ESKVEIFSDLPFRAIELEF 53
OY 99 QWNLRKRPDPCFKNYHDLOKIDONNKITSISYAFGLNSLTKLYSHNRITFLKPGV 158
DB 54 VLTKRVYQKGFSGFGLKEIKESQNDVLE-----IEADV 90
OY 159 FEDLHREMLIIE-NHLSRISPPFFYGLNSLILVLMNNVLRPLDKPCLQHMPLRL-- 215
DB 91 FSNLPKLEHRIEIKANNMLTYNPEAFQNLPLNQLYILNSTGTGIKHLPD-----VHKIHSL 144

```

```

OY 216 ---MDEEGN-HIHNLRLNLFISCS-NLTVLMRKNIHNLNENTFADLOKLELDG-S 269
DB 145 QKVLIDIONINHTIENRSFVGLSFESVITLMKNGIQEIHNCFAFNQO-LDEINLSDN 203
OY 270 NKIENLPPLIFKDKELISQNLMSYNPDKIQANOFDYLKLSLSEIG- 319
DB 204 NNLEELPNDVFGASGPYILDISRTRHSLPSYGLENKILRARSTYLNKKLPTEKLV 263
OY 320 -----ISNIQRM--FRPLMNS----- 335
DB 264 LMEASLTPSHOCAPFANMRROISELHPICNKSILRQEVDMYOTRGORSSLAEDNESSYS 323
OY 336 ---HYEFKFOY--CGYAPHVRSCKPMDGTSLENLASTIORFVWVAVTCGNIF 390
DB 324 RGFDMTYEFEDYDLCNEVVDV-TCSPPKPDARPCDDINGNILRLIMFISLATGNI 382
OY 391 VICMRPIRSENKLYAMSII-SLOCADCLMGIVLFVIGGFDLKRGEYNKRAOLMEST 448
DB 383 VLVI--LFTSQYKLIVPRFLMCLNLAFAADLCIGIYLLILASVMDIHKSQYHNVAIDMOTGA 440
OY 449 HCQLVGLAIISTEVSULLFPLETKYICVYPRCYRPKCR-----TIVLILWIT 503
DB 441 GCDAGFTYVASELSVYTLTAILERWHITTHAQL-----DCKYQLRHAASVWGMIF 496
OY 504 GVIYAFILPSNKEPKNYGTNGVCFPHSEDESIGAIYSVALFGLINLAFFIYES 563
DB 497 AFNALFPI-----FGISSYMKVSTCLPM---DIDPLSQLYMSL-LVNLVIAFVITGC 548
OY 564 YGSMFYVHQSATYATERNQVKEMILAKREFYFTDALCWDPIFYVFKLSLQVEIP 623
DB 549 YHIIYLRNPNIVAS-----SSDTRIARKRMALIFDFCLMADISPAISAKVPLI 602
OY 624 GTITSWVYFIL--PINSALNPIYTLTPRK 654
DB 603 -TVSKAKILLVLEPHINSCANDELYAIFTKNR 634

RESULT 12
FSHR_MOUSE STANDARD; PRT; 692 AA.
AC P35378; Q9QWV8; Q9DAC2;
DT 01-JUN-1994 (Rel. 29, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Follicle stimulating hormone receptor precursor (FSH-R) (Follicotropin receptor).
GN FSHR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RC SEQUENCE FROM N.A.
RC STRAIN=129/SV; TISSUE=Testis;
RC Tena-Semper M., Manna P.R., Huhtaniemi I.T.;
RT "Molecular cloning of the mouse follicle stimulating hormone receptor complementary deoxyribonucleic acid: functional expression of alternatively spliced variants and receptor inactivation by a C566T transition in exon 7 of the coding sequence."
RT submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RC MEDLINE=21085660; PubMed=11217851;
RX Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arikawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.,
RA Saito T., Okazaki Y., Gojibori T., Bono H., Kasukawa T., Saito R.,
RA Kadoya K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Piesle G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,

```

RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaudo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Welter C., Whitaker C., Wilming L.,
 RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
 RA Hayashizaki Y.,
 RA "Functional annotation of a full-length mouse cDNA collection.",
 RL Nature 409:685-690(2001).
 RN [3]
 RP SEQUENCE OF 1-51 FROM N.A.
 RX MEDLINE=93093308; PubMed=1459341;
 RA Hultin I.T., Eskola V., Pakarinen P., Matikainen T.,
 RA Sprengel R.,
 RT "The murine luteinizing hormone and follicle-stimulating hormone
 RT receptor genes: transcription initiation sites, putative promoter
 RT sequences and promoter activity.",
 RL Mol. Cell. Endocrinol. 88:55-66(1992).
 CC -1- FUNCTION: RECEPTOR FOR FOLLICLE STIMULATING HORMONE. THE ACTIVITY
 CC OF THIS RECEPTOR IS MEDIATED BY G PROTEINS WHICH ACTIVATE
 CC ADENYLATE CYCLASE.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
 CC FSH/LSH/TSH SUBFAMILY.
 CC -1- SIMILARITY: CONTAINS 6 LEUCINE-RICH REPEATS (LRR).
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF095642; AAC67559.1; -;
 DR EMBL: AK016635; BAB30351.1; -;
 DR EMBL: S49632; AAB24401.1; -;
 DR EMBL: M87570; AAA37641.1; -;
 DR GCRDP: GCR.0304; -;
 DR MGD: MGI:95583; Fshr.
 DR InterPro: IPR000276; GPCR_Rhodopsn.
 DR InterPro: IPR001611; LRR.
 DR InterPro: IPR000372; LRR_Nterm.
 DR Pfam: PF00001; 7tm_1; 1.
 DR Pfam: PF00560; LRR; 3.
 DR Pfam: PF01462; LRRNT; 1.
 DR PRINTS: PR00237; GPCR_RHODOPSIN.
 DR SMART: SM00013; LRRNT; 1.
 DR PROSITE: PS00237; G-PROTEIN_RECP_F1_1; 1.
 DR PROSITE: PS00262; G-PROTEIN_RECP_F2_1; 1.
 KM G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;
 KM Phosphorylation; Repeat; Leucine-rich repeat.
 FT SIGNAL 1 17
 FT CHAIN 18 692
 FT DOMAIN 18 365
 FT TRANSSEM 366 386
 FT DOMAIN 387 397
 FT TRANSSEM 398 420
 FT DOMAIN 421 442
 FT TRANSSEM 443 464
 FT DOMAIN 465 507
 FT TRANSSEM 508 527
 FT DOMAIN 528 549
 FT TRANSSEM 550 572
 FT DOMAIN 573 596
 FT TRANSSEM 597 607
 FT DOMAIN 608 629
 FT TRANSSEM 630 692
 FT DOMAIN 630 692
 FT REPEAT 44 68
 FT REPEAT 69 93

FT REPEAT 119 143
 FT REPEAT 170 192
 FT REPEAT 193 216
 FT REPEAT 218 240
 FT REPEAT 441 516
 FT CARBOHYD 191 191
 FT CARBOHYD 199 199
 FT CARBOHYD 293 293
 FT CONFLICT 436 436
 SQ SEQUENCE 692 AA: 77769 MW: 4B57229180563A44 CRC64;
 Query Match 12.4%; Score 475.5; DB 1; Length 692;
 Best Local Similarity 23.5%; Pred. No. 8,5e-22;
 Matches 162; Conservative 125; Mismatches 265; Indels 137; Gaps 25;
 QY 35 GNTTKLPQLHNGVDDCGNQADENCVVYLCOGMSLPGELDMMKRFVSVSSTVT 94
 DB 13 GSGGCHHMLCHCSN-----RVFLCODSKVTEIP-----PDLPRNAI 49
 QY 95 AMSLOWNLIRKLPPDCFRNYHDLQKDL-ONNKLTISIVAFRGNSLTKYLSH-NRIT 152
 DB 50 ELRFVLTKRYTPKGSFSGFDELEKIEISQNDVLEVADYFSLPNIHEIRKANNL 109
 QY 153 FLKPGVEDLRLMLIEDNHLNRISPPPEYGLNSL-ILLVMMNVYTRLPDKPLCOH 210
 DB 110 YINPEARQNLPSLKYLLISNTGIKHL-PAFHKIQSLQKVLDDIQDN----- 155
 QY 211 MPRLMWLDLBSNHHNRLNLTFCSS-NLYLVKRRKKNHNLNENTRAPLOKDELDG- 268
 DB 156 -----NHIIARNSFMGLSFESVYLMANKNGIOGRHNCANGTO-IDELNLSID 202
 QY 269 SNKIENPLPIFKD-----LKEISQNLNLYNDYIOQANQPDYVTKLSLS-L 315
 DB 203 NNNEELPDYFQASGPVYLDISRTKYSLPNIENLEKRLKARSRYRLKPLSLDKFV 262
 QY 316 EGIEIS-----NIQRMFRPLMNL----- 334
 DB 263 MLIPASLTPSPHCCAFANWRQTSSELPHCNKSSISRODIDDMTPGDQRVSLVDEPSYG 322
 QY 335 --SHIYFKKPY--CGAAPHYRSGKPRNTDGISSLENLASSIQVFWVYVAVCFENIF 390
 DB 323 KGSOMLYSEFDYDCNEFVDY-TCSPKPDFAFNPCEDIMGVILVILWFISLAIITNTT 381
 QY 391 VICRPIYRSENKLYAMSI--SLCCADLMGIYLFYIGFDELFKRGYNHQAQMEST 448
 DB 382 VLWY--LFTSQYKLTVRFLMNCNLAFLADLCIGYLLILLASVDHTKSOYHVAIDMPTGA 439
 QY 449 HQQLVGSIALISTEVSVLLTFLTELYKVICIYVPRC-VPRGKRTTVLLIWIQFTIV 507
 DB 440 GCDAAGFETVAFASLTYTLAATILFERMHTTHAMQLECKVQLCHAASIMVLGWAFAPAA 499
 QY 508 AFPLSKKEPKKNYNGVNGCFPLHSDTESIGAOISVAFIIGINLAFFIIFYSGSM 567
 DB 500 ALFPI---FGISSYKVSICLPW---DIDPSQLTYMAL-LVLNMALEFVIGCYTHI 551
 QY 568 FYVSHOSAITATEIRNOVKKEMILAKRFFVFETDACPWFIVFKLSLOYEIPETIT 627
 DB 552 YLYVRNNIVSS-----SNDTKAKRMATLIFDFLCMPILFEAFASLSKVPIL-TVS 604
 QY 628 SWVVIPL--PINSALNPILYTLTPPK 654
 DB 605 KAKILVLEFPINSCANPFLAIFTKMR 633
 RESULT 13
 FSHR_EODAS STANDARD; PRT; 687 AA.
 AC 095179;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Follicle stimulating hormone receptor precursor (FSH-R) (foliitropin

DE receptor).
 GN FSHR.
 OS Equus asinus (Donkey).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
 OX NCBI_TaxID=9793;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Testis;
 RX MEDLINE=97389913; PubMed=9195473;
 RA Richard F., Martinat N., Remy J.-J., Salese R., Combarnous Y.;
 RT "Cloning, sequencing and in vitro functional expression of
 RT recombinant donkey follicle-stimulating hormone receptor: a new
 RT insight into the binding specificity of gonadotrophin receptors.";
 RL J. Mol. Endocrinol. 18:193-202(1997)
 CC -1- FUNCTION: RECEPTOR FOR FOLLICLE STIMULATING HORMONE. THE ACTIVITY
 CC OF THIS RECEPTOR IS MEDIATED BY G PROTEINS WHICH ACTIVATE
 CC ADENYLATE CYCLASE.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
 CC FSH/LSH/TSH SUBFAMILY.
 CC -1- SIMILARITY: CONTAINS 6 LECINE-RICH REPEATS (LRR).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U73659; AAB18245.1; -.
 DR HSSP: P23945; 1XUN.
 DR GCRdb: GCR_1116; -.
 DR InterPro: IPR000276; GPCR_Rhodpsn.
 DR InterPro: IPR001611; LRR.
 DR InterPro: IPR00372; LRR_Nterm.
 DR Pfam: PF00001; 7tm_1; 1.
 DR Pfam: PF00560; LRR; 4.
 DR Pfam: PF01462; LRRNT; 1.
 DR PRINTS: PR00237; GPCR_RHODPSN.
 DR SMART: SM00013; LRRNT; 1.
 DR PROSITE: PS00237; G_PROTEIN_RECPT_FL1; 1.
 DR PROSITE: PS50262; G_PROTEIN_RECPT_FL2; 1.
 KW G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;
 KW Phosphorylation; Repeat; Leucine-rich repeat.
 FT SIGNAL 1 17
 FT CHAIN 18 687
 FT DOMAIN 18 358
 FT TRANSSEM 359 379
 FT DOMAIN 380 390
 FT TRANSSEM 391 413
 FT DOMAIN 414 435
 FT TRANSSEM 436 457
 FT DOMAIN 458 477
 FT TRANSSEM 478 500
 FT DOMAIN 501 520
 FT TRANSSEM 521 542
 FT DOMAIN 543 565
 FT TRANSSEM 566 589
 FT DOMAIN 590 600
 FT TRANSSEM 601 622
 FT DOMAIN 623 687
 FT REPEAT 44 68
 FT REPEAT 69 93
 FT REPEAT 119 143
 FT REPEAT 170 192
 FT REPEAT 193 216
 FT REPEAT 218 240
 FT DISULFID 434 509
 FT CARBOHYD 191 191
 FT CARBOHYD 199 199
 FT CARBOHYD 293 293

SQ SEQUENCE 687 AA; 76937 MW; FC3AF0B5531DA9A CRC64;
 Query Match 12.4%; Score 474.5; DB 1; Length 687;
 Best Local Similarity 24.1%; Pred. No. 9,7e-22;
 Matches 161; Conservative 126; Mismatches 255; Indels 127; Gaps 24;
 QY 53 CGNQADENCVYVLCQCMSPGLGELDMKPKPTSVS-VSSNVTAMSLQMNIRKPPPCF 111
 DB 18 CHQVCHYSNRFVLCQ-----ESKVEYTSDDLPNNLELREVLVTLRYPKGAF 66
 QY 112 KNYHDLQKLDLQNNKITSISYAFRGLNSLTKLYLSHNRITFLKPGVEDLRLLEWITE 171
 DB 67 SGFGDLKKEIEISQNDVLEV-----IEANVSNPKLKEIRIE 103
 QY 172 D-NHLSRISPPTEFGNSLILVLMNNVLTLPDKPQLQHPRLWDLQGN-HIHNLRN 229
 DB 104 KANNLLYIDHDAFQNLPMIQLYLLISNTGICKLPVAKKIOSLOKV-LTDIDQNNIHIVER 162
 QY 230 LPTFISCS-NLAVLVYRRKKINHLNENTFAPLOKIDELDG-SNKIENLPPLFKDLKELS 287
 DB 163 NSFGLSESMILKLSKNGIDEIFNCAFNQTO-LDELNLSNNNLEELPNDVFGASGPV 221
 QY 288 QNLSTYNEIQRIOANQFYLVLKLSLEGI-EISNIOQRMRPLMNLSHY----- 338
 DB 222 ILDISGTRHSLPNVGLNKLKRLARSITYNLKLPSLEK--FVALMEASLTPYSHCAFA 279
 QY 339 -----FKKQY--CGYAPHVNSC 354
 DB 280 NWRQOTSELOTCNKSILQOEVDMTQARGERYSLAEDDESMMYSSEFDYDLCNEVDV-TC 338
 QY 355 KPNIDGISLENLALSIQRYFVWVAVTCGNFVYCMRPYISENKLAMSLI--SL 412
 DB 339 SPKPDANPCEDIMGYDLRLVLMFISTLATIGNIIVLT--LITSQKLVYPRRLKCNL 396
 QY 413 CCADCLMGIVYLVIGGFDLKEGEYKKAQOLMESTHCQVSLALISTEVSALLTFELT 472
 DB 397 AFADLCIGIYLLILASVVDIHRKSQYHNVAIDMQTAGDAGGFVFPSELSTVYLTAIT 456
 QY 473 LEKYCIYPPRCVRRPKCR-----TIVLLIIVITGFIYAFIPLSNKEFFKNYGTNGV 527
 DB 457 LERWHTITHAMQL-----ECKVOLRHAASVMYGVFGVGVGLPI-----FGISTYMKVSI 508
 QY 528 CEPHLSEPTESGAIQYSAVATFLGINLAFTIIVSYSGSMFYSVQASATAEINQYKK 587
 DB 509 CLPM--DIDPSLSQLYMSL-LVLNVAIFVYICGCTHYHITLVNPNIVS-----SS 558
 QY 588 EMILAKRFFEYFTDALCMIPFVVKFSLQVEIPGTTSMVVFIL--PINSALNPL 645
 DB 559 DTKIKRKGILITFDPLCMAPISFSGISASLKVALLI-IVSKSKILLVLFPYINSCANFL 617
 QY 646 YTLTRTPRK 654
 DB 618 YAIPTKKNR 626
 RESULT 14
 LSHR_MOUSE STANDARD; PRT; 700 AA.
 AC P30730;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Lutropin-choriogonadotropin hormone receptor precursor (LH/CG-R)
 GN (LSH-R) (Luteinizing hormone receptor).
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 MEDLINE=92165799; PubMed=1311310;

[illegible]

RP SEQUENCE FROM N.A.
RA MEDLINE=89332512; PubMed=2502842;
RX McFarland K.C., Sprengel R., Phillips H.S., Koehler M.,
RA Rosenblat N., Nikolics K., Segaloff D.L., Seeburg P.H.;
RT "Lutropin-choriogonadotropin receptor: an unusual member of the G
RL protein-coupled receptor family."; Science 245:494-499(1989).
[2]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Ovary;
RX MEDLINE=92347604; PubMed=1353463;
RA Aatsinki J.T., Pietila E.M., Lakkakorpi J.T., Rajanemi H.J.;
RT "Expression of the LH/CG receptor gene in rat ovarian tissue is
RT regulated by an extensive alternative splicing of the primary
RT transcript"; Mol. Cell. Endocrinol. 84:127-135(1992).
[3]
RP SEQUENCE FROM N.A.
RX MEDLINE=91209270; PubMed=2019252;
RA Koo Y.B., Slaughter R.G., Ji T.H.;
RT "Structure of the lutelizing hormone receptor gene and multiple
RT exons of the coding sequence."; Endocrinology 128:2297-2308(1991).
[4]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX MEDLINE=91006819; PubMed=1976554;
RA Bernard M.P., Myers R.V., Moyle W.R.;
RT "Cloning of rat lutropin (LH) receptor analogs lacking the soybean
RT lectin domain."; Mol. Cell. Endocrinol. 71:R19-R23(1990).
[5]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX MEDLINE=91126285; PubMed=2281186;
RA Segaloff D.L., Sprengel R., Nikolics K., Ascoli M.;
RT "Structure of the lutropin/choriogonadotropin receptor."; Recent Prog. Horm. Res. 46:261-303(1990).
[6]
RP SEQUENCE OF 295-700 FROM N.A.
RX MEDLINE=91060531; PubMed=2174034;
RA Tsai-Morris C.H., Buckzo E., Wang W., Dufau M.L.;
RT "Intonic nature of the rat lutelizing hormone receptor gene defines
RT a soluble receptor subspecies with hormone binding activity."; J. Biol. Chem. 265:19385-19388(1990).
[7]
RP SEQUENCE OF 27-37.
RX MEDLINE=89174723; PubMed=2925659;
RA Roche P.C., Ryan R.J.;
RT "Purification, characterization, and amino-terminal sequence of rat
RT ovarian receptor for lutelizing hormone/human chorogonadotropin."; J. Biol. Chem. 264:4636-4641(1989).
[8]
RP MUTAGENESIS.
RX MEDLINE=91332007; PubMed=1714448;
RA Ji I., Ji T.H.;
RT "Asp383 in the second transmembrane domain of the lutropin receptor
RT is important for high affinity hormone binding and cAMP production."; J. Biol. Chem. 266:14953-14957(1991).
-1- FUNCTION: RECEPTOR FOR LUTROPIN-CHORIOGONADOTROPIC HORMONE.
CC THE ACTIVITY OF THIS RECEPTOR IS MEDIATED BY G PROTEINS WHICH
CC ACTIVATE ADENYLYLATE CYCLASE.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- ALTERNATIVE PRODUCTS: AT LEAST 11 ISOFORMS WHICH DIFFER IN
CC SUBCELLULAR LOCATION ARE PRODUCED BY ALTERNATIVE SPLICING
CC OF THE SAME GENE.
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC FSH/LSH/TSH SUBFAMILY.
CC -1- SIMILARITY: CONTAINS 7 LEUCINE-RICH REPEATS (LRR).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

CC entitles requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M26199; AAA41528.1; -
DR EMBL; M61212; AAA41527.1; -
DR EMBL; M61211; AAA41527.1; JOINED.
DR EMBL; S40803; AAB22680.1; -
DR EMBL; S40787; AAB22680.1; JOINED.
DR EMBL; S40904; AAB22680.1; JOINED.
DR EMBL; S40905; AAB22680.1; JOINED.
DR EMBL; S40905; AAB22680.1; JOINED.
DR EMBL; S40909; AAB22680.1; JOINED.
DR EMBL; S40918; AAB22680.1; JOINED.
DR EMBL; S40920; AAB22680.1; JOINED.
DR EMBL; S40795; AAB22680.1; JOINED.
DR EMBL; S40798; AAB22680.1; JOINED.
DR EMBL; S40795; AAB22681.1; -
DR EMBL; S40787; AAB22681.1; JOINED.
DR EMBL; S40903; AAB22681.1; JOINED.
DR EMBL; S40904; AAB22681.1; JOINED.
DR EMBL; S40905; AAB22681.1; JOINED.
DR EMBL; S40907; AAB22681.1; JOINED.
DR EMBL; S40918; AAB22681.1; JOINED.
DR EMBL; S40920; AAB22681.1; JOINED.
DR EMBL; S40803; AAB22682.2; -
DR EMBL; S40787; AAB22682.2; JOINED.
DR EMBL; S40903; AAB22682.2; JOINED.
DR EMBL; S40909; AAB22682.2; JOINED.
DR EMBL; S40918; AAB22682.2; JOINED.
DR EMBL; S40920; AAB22682.2; JOINED.
DR EMBL; S40795; AAB22683.1; JOINED.
DR EMBL; S40798; AAB22683.1; JOINED.
DR EMBL; S40904; AAB22683.1; JOINED.
DR EMBL; S40905; AAB22683.1; JOINED.
DR EMBL; S40907; AAB22683.1; JOINED.
DR EMBL; S40909; AAB22683.1; JOINED.
DR EMBL; S40918; AAB22683.1; JOINED.
DR EMBL; S40920; AAB22683.1; JOINED.
DR EMBL; S40795; AAB22683.1; JOINED.
DR EMBL; S40798; AAB22683.1; JOINED.
DR EMBL; S40803; AAB22684.2; -
DR EMBL; S40787; AAB22684.2; JOINED.
DR EMBL; S40903; AAB22684.2; JOINED.
DR EMBL; S40904; AAB22684.2; JOINED.
DR EMBL; S40905; AAB22684.2; JOINED.
DR EMBL; S40909; AAB22684.2; JOINED.
DR EMBL; S40918; AAB22684.2; JOINED.
DR EMBL; S40920; AAB22684.2; JOINED.
DR EMBL; S40795; AAB22684.2; JOINED.
DR EMBL; S40798; AAB22684.2; JOINED.
DR EMBL; M68928; AAA41529.1; -
DR EMBL; M68917; AAA41529.1; JOINED.
DR EMBL; M68918; AAA41529.1; JOINED.
DR EMBL; M68919; AAA41529.1; JOINED.
DR EMBL; M68920; AAA41529.1; JOINED.
DR EMBL; M68921; AAA41529.1; JOINED.
DR EMBL; M68922; AAA41529.1; JOINED.
DR EMBL; M68923; AAA41529.1; JOINED.
DR EMBL; M68925; AAA41529.1; JOINED.
DR EMBL; M68926; AAA41529.1; JOINED.
DR EMBL; M68927; AAA41529.1; JOINED.
DR PIR; A32460; A32460.
DR PIR; A41343; A41343.
DR HSP; P22886; 1IUT.
DR GCRDb; GCR_0138; -
DR GCRDb; GCR_0139; -
DR GCRDb; GCR_0262; -

```

DR GCRDB; GCR_0612; -
DR GCRDB; GCR_0613; -
DR GCRDB; GCR_0614; -
DR InterPro; IPR000276; GPCR_Rhodopsn.
DR InterPro; IPR001611; LRR.
DR Pfam; PF00001; 7tm_1; 1.
DR Pfam; PF00560; LRR; 2.
DR PRINTS; PRO0373; GLYCHORMONER.
DR PRINTS; PRO1144; LSHRECEPTOR.
DR SMART; SM00013; LRNT; 1.
DR PROSITE; PS00237; G-PROTEIN_RECEP_FL_1; 1.
DR PROSITE; PS00262; G-PROTEIN_RECEP_FL_2; 1.
DR G-protein coupled receptor; Transmembrane; signal;
DR Phosphorylation; Repeat; leucine-rich repeat; Alternative splicing.
FT STGNL 1 26
FT CHAIN 27 700
FT DOMAIN 27 362
FT TRANSMEM 363 390
FT DOMAIN 391 399
FT TRANSMEM 400 422
FT DOMAIN 423 443
FT TRANSMEM 444 466
FT DOMAIN 467 486
FT TRANSMEM 487 509
FT DOMAIN 510 529
FT TRANSMEM 530 551
FT DOMAIN 552 574
FT TRANSMEM 575 598
FT DOMAIN 599 609
FT TRANSMEM 610 631
FT DOMAIN 632 700
FT REPEAT 52 75
FT REPEAT 126 150
FT REPEAT 152 175
FT REPEAT 176 200
FT REPEAT 202 224
FT REPEAT 225 248
FT REPEAT 250 271
FT DISULFID 443 518
BY SIMILARITY.

```

Query Match 12.3%; Score 470.5; DB 1; Length 700;
 Best Local Similarity 22.8%; Pred. No. 1,7e-21;
 Matches 149; Conservative 135; Mismatches 233; Indels 137; Gaps 23;

```

QY 87 PVSASNTAMSLOMNLRLKLPDPCFKNYHDLQKIDLQNNKITSISYARGLNSLTKLY 145
DB 49 GPPRAGLARLSLYLPVKYIPSAQAFGLNEVYKIEISQDSLRIEANAFAFDNLNLSEL 108
QY 146 LSHNR-ITFLKPGVEEDLHLREMLIEDN-----HLSRISPTFFYGLNSLILVMNNV 198
DB 109 IQNTKNLITIEPGAFTLPLRLKYISICNTGRTLPDYTKISSSEF-----NFILEICDNLH 164
QY 199 LTRLDPRLCQHMRLHMLDEGNHINLRLNLFISCSNLTIVMRKNKINHLNENTFAP 258
DB 165 ITTIPGNAFOGNNESVTLYKYGNGFEVQSHAF-----NGTTLISLE-----LKENIY-- 213
QY 259 LQKIDE-----LDGSKNTENLPLRIKDLKELSQLNSYNPIQKIQANOPDY 307
DB 214 LEKHSGAFOGANGPSIILDISITKLALPS-----HGLESIQF-----L 252
QY 308 VKLKSLSLEGLEISNIQORMERPLM-----NLSHYFKKQFQ--- 343
DB 253 IALSSYSILKTLF-----SKEKFTSLVATLYTPSHCAFRNLPKKEONFSFISFENSKOC 308
QY 344 -----YCGYAPHVRSCKPNTDGISSLENLNASIIQ 373
DB 309 ESTYRKADNETLISAIFENNELSGWDYDGC-SPKTIQCAPEDPAFNPCEIDIMGTAFL 366
QY 374 RVEYVWVSAVTCGNIIVICMRPIYRSENKLYAMSIISLCCADCLMGIVLFVIGGFDLKF 433
DB 367 RVLIMLIIILAIIGNLVLELVLLTSRYKLTVPRLMCNLSFADFQCMGLAYILLIASVDSQT 426

```

```

QY 434 RGEYKHAQIMMESTHCOLVSLAISTEVSULLLFLLEKYICIVYPPRCVPGKCR- 492
DB 427 KGOYVNAIDMONGSGGGAAGFTVFASFELSVYTLITLERHTITYANQDLKRLRH 486
QY 493 TITVLLIWTGFIYAFIPISNKEEFKNYGTNGVCPFLHSEDTESIGAOIYSAIFLGI 552
DB 487 AIPIMLGMLFSTLIATMPLVG---ISNMYKVS-ICLPM--DVESTLSQVYLISILI-L 538
QY 553 NIAFIIIVSYSGMFSYVHQSATLATEINQVKKEMILAKRFFIYFTDALCMPIFVY 612
DB 539 NVAFVVIACYIRIYPAVONPELTAP-----NKDKIKAKMAILITFTCAPISSEF 592
QY 613 KFLSLQVEIPIGTTSWVIFIL--PINSALNPILYTLTRPERE---MIHRE 660
DB 593 AISAAPKVPPLI-IVTNSKILLVLEFPYNSCANPFLVAFITKAFQDRDLILLSR 645

```

Search completed: September 5, 2002, 09:04:06
 Job time: 2325 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2002, 08:24:36 : Search time 85.41 Seconds

(without alignments)
1462.385 Million cell updates/sec

Title: US-09-647-067-8

Perfect score: 3834
Sequence: 1 MTSGSVFFILLIFKXFSHG.....FTYPCMSLISQSTRINSYS 722

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

- 1: SPREMBL_19:*
- 2: sp.archaea:*
- 3: sp.bacteria:*
- 4: sp.fungi:*
- 5: sp.human:*
- 6: sp.invertebrate:*
- 7: sp.mhcf:*
- 8: sp.organelle:*
- 9: sp.phage:*
- 10: sp.plant:*
- 11: sp.protist:*
- 12: sp.virus:*
- 13: sp.vertebrate:*
- 14: sp.unclassified:*
- 15: sp.rvivirus:*
- 16: sp.bacteriap:*
- 17: sp.archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3687.5	96.2	757	4 Q9HBX9	Q9hbxb9 homo sapien
2	1891.5	49.3	737	11 Q91ZT5	Q91z25 mus musculu
3	599.5	15.6	359	5 Q9VBP0	Q9vbp0 drosophila
4	576	15.0	1012	5 Q95YI6	Q95y16 asterina pe
5	576	15.0	1280	5 Q95YI7	Q95y17 asterina pe
6	496	12.9	1050	5 Q9BN18	Q9bn18 drosophila
7	494	12.9	701	13 Q9DGC6	Q9dgc6 oreochromis
8	492.5	12.8	1300	5 Q9NMD6	Q9nmd6 oreochromis
9	490	12.8	907	11 Q9Z1P4	Q9z1p4 mus musculu
10	486.5	12.7	1360	5 Q9ND11	Q9nd11 drosophila
11	483.5	12.6	658	13 Q9PYN9	Q9pyn9 oncorhynch
12	481	12.5	693	13 Q9DGC5	Q9dgc5 oreochromis
13	479.5	12.5	696	13 Q9DGF5	Q9dgt5 cynops pyrr
14	478	12.5	662	13 Q9PW16	Q9pw16 clarias gar
15	477.5	12.5	334	5 Q9VYG0	Q9vyg0 drosophila
16	476.5	12.4	907	4 Q75473	Q75473 homo sapien

17	476.5	12.4	951	11 Q9Z2H4	Q9z2h4 rattus norv
18	475	12.4	699	4 Q15996	Q15996 homo sapien
19	472.5	12.3	688	11 Q64183	Q64183 rattus sp.
20	470.5	12.3	907	4 Q9UP75	Q9up75 homo sapien
21	470	12.3	662	13 Q98R84	Q98r84 ictalurus p
22	469	12.2	701	4 Q14751	Q14751 homo sapien
23	467	12.2	951	4 Q9BXD1	Q9bxd1 homo sapien
24	464	12.1	951	4 Q9NYD1	Q9nyd1 homo sapien
25	453	11.8	764	6 Q9BG56	Q9bg56 sus scrofa
26	443	11.6	696	13 Q98T85	Q98t85 ictalurus p
27	443	11.6	793	13 Q91948	Q91948 oncorhynch
28	441.5	11.5	764	11 Q9D697	Q9d697 mus musculu
29	439.5	11.5	763	6 Q9BGM4	Q9bgm4 felis silve
30	439	11.5	739	6 Q9BG55	Q9bg55 sus scrofa
31	428	11.2	814	13 Q91949	Q91949 oncorhynch
32	427.5	11.2	778	13 Q98TF4	Q98tf4 oreochromis
33	425	11.1	779	13 Q91BN7	Q91bn7 morone saxa
34	416.5	10.9	928	4 Q9BYD7	Q9byd7 homo sapien
35	403	10.5	601	13 Q42500	Q42500 melleagris g
36	402	10.5	829	5 Q9VEG4	Q9veg4 drosophila
37	402	10.5	831	5 Q94979	Q94979 drosophila
38	400.5	10.4	673	13 Q90WP8	Q90wp8 podarcis si
39	394	10.3	828	4 Q9HBX8	Q9hbxb8 homo sapien
40	389	10.1	533	4 Q9P1V4	Q9p1v4 homo sapien
41	387.5	10.1	724	13 Q9PVP0	Q9pvp0 oncorhynch
42	385	10.0	470	4 Q43200	Q43200 homo sapien
43	373.5	9.7	929	5 Q18759	Q18759 caenorhabdl
44	353	9.2	293	6 Q95MF7	Q95mf7 canis famli
45	340.5	8.9	410	4 Q16225	Q16225 homo sapien

ALIGNMENTS

RESULT	ID	Q9HBX9	PRELIMINARY:	PRT:	757 AA.
AC	Q9HBX9				
DT	01-MAR-2001	(TREMBLREL. 16, Created)			
DT	01-MAR-2001	(TREMBLREL. 16, Last sequence update)			
DT	01-DEC-2001	(TREMBLREL. 19, Last annotation update)			
DE	LEUCINE-RICH REPEAT-CONTAINING G PROTEIN-COUPLED RECEPTOR 7.				
GN	LGR7.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=20388592; PubMed=10935549;				
RA	Hsu S.Y., Kudo M., Chen T., Nakabayashi K., Bhalla A.,				
RA	van der Spek P.J., Van Duin M., Hsueh A.J.;				
RT	"The three subfamilies of leucine-rich repeat-containing G protein-				
RT	coupled receptors (LGR): identification of LGR6 and LGR7 and the				
RT	signaling mechanism for LGR7."				
RL	Mol. Endocrinol. 14:1257-1271(2000).				
DR	EMBL; AF190500; AAC17167.1; -				
DR	HSSP; P01130; 1LDR.				
DR	InterPro; IPR000276; GPCR_Rhodpsn.				
DR	InterPro; IPR002172; LDL_recept_A.				
DR	InterPro; IPR001611; LRR.				
DR	InterPro; IPR000372; LRR_Nterm.				
DR	InterPro; IPR003592; LRR_out.				
DR	InterPro; IPR003591; LRR_Typ.				
DR	Pfam; PF00001; 7tm_1; 1.				
DR	Pfam; PF00057; 1dl_recept_a; 1.				
DR	Pfam; PF00560; LRR; 8.				
DR	PRINTS; PR00237; GPCRHHODPSN.				
DR	SMART; SM00192; LDLa; 1.				
DR	SMART; SM00370; LRR; 7.				
DR	SMART; SM00013; LRRNT; 1.				
DR	SMART; SM00369; LRR_Typ; 10.				
DR	PROSITE; PS50262; G_PROTEIN_RECPT_FL_2; 1.				

DR PROSITE: PS01209; LDLRA_1; 1.
 DR PROSITE: PS50068; LDLRA_2; 1.
 KM Glycoprotein; Receptor.
 SQ SEQUENCE 757 AA; 86992 MW; 8079EBDBF3A3EF21 CRC64;

Query Match 96.2%; Score 3687.5; DB 4; Length 757;
 Best Local Similarity 93.0%; Pred. No. 2.4e-268;
 Matches 707; Conservative 2; Mismatches 10; Indels 41; Gaps 3;

QY 1 MMSGVFYLLIFGKYFSGGGGADYKCSGYPCGNTKCLPQLLHCNGVDCCGADDED 60
 |||||
 DB 1 MMSGVFYLLIFGKYFSGGGGADYKCSGYPCGNTKCLPQLLHCNGVDCCGADDED 60
 |||||
 QY 61 NC-----VVVLCQMSLPGLELDMKP- 82
 |||||
 DB 61 NCQDNNNGMOPDKYFASYYKMTSOYFPEAEPECELVGSPVQCLQ---GLELDCDEFN 117
 |||||
 QY 83 FTSVPSVSSNVTAMSIQWNLIRKLPDCKRKNYHDLOKLDLONNKITSISYAFRGINSLT 142
 :
 DB 118 LRAVPSSNVTAMSIQWNLIRKLPDCKRKNYHDLOKLDLONNKITSISYAFRGINSLT 177
 |||||
 QY 143 KLYLSNRTPELPGVEEDLHREMLITEDNHLSTRISPPFVGLNSLILLVLMNNVLTFL 202
 |||||
 DB 178 KLYLSNRTPELPGVEEDLHREMLITEDNHLSTRISPPFVGLNSLILLVLMNNVLTFL 237
 |||||
 QY 203 PDKPLCQHPRLHMLDEGNHINHLNLTFTISCSNLTVLVMRKNKINHLENTFAPLQKI 262
 |||||
 DB 238 PDKPLCQHPRLHMLDEGNHINHLNLTFTISCSNLTVLVMRKNKINHLENTFAPLQKI 297
 |||||
 QY 263 DELDGSNKIENLPILFDKLKLSQNLNSYNPQIKQANQFYLKKSLSLEGIEISN 322
 |||||
 DB 298 DELDGSNKIENLPILFDKLKLSQNLNSYNPQIKQANQFYLKKSLSLEGIEISN 357
 |||||
 QY 323 IOORMRPLMNSHIEFKKFOYCYAPHYRSCPNPDGSIENLNASIQRFVWVSA 382
 |||||
 DB 358 IOORMRPLMNSHIEFKKFOYCYAPHYRSCPNPDGSIENLNASIQRFVWVSA 417
 |||||
 QY 383 VYCFGNIFVYCMRPYIRSEKRLYAMSIIISLCADCLMGITLFIYIGGFDLKFGEYKHAQ 442
 |||||
 DB 418 VYCFGNIFVYCMRPYIRSEKRLYAMSIIISLCADCLMGITLFIYIGGFDLKFGEYKHAQ 477
 |||||
 QY 443 LMESEHCOLVGLSLSTFVSALLFTFLTEKYICIVPERCVRGKCTIVLLIWI 502
 |||||
 DB 478 LMESEHCOLVGLSLSTFVSALLFTFLTEKYICIVPERCVRGKCTIVLLIWI 537
 |||||
 QY 503 TGRIVAFIPLSNKEFFKNYGTNGVCFPLHSEPTESIGAOIYVAIFLGINLAFFIIV 562
 |||||
 DB 538 TGRIVAFIPLSNKEFFKNYGTNGVCFPLHSEPTESIGAOIYVAIFLGINLAFFIIV 597
 |||||
 QY 563 SYGSMFYVHQAATATEIRNOYKKEMLAKRPFVFTDALCWPILFVVKFSLQVEI 622
 |||||
 DB 598 SYGSMFYVHQAATATEIRNOYKKEMLAKRPFVFTDALCWPILFVVKFSLQVEI 657
 |||||
 QY 623 PGITTSWVVFILPINSALNPILYTLTRPKEIHRFWMYNRKSMDSKGGKTAPSF 682
 |||||
 DB 658 PGITTSWVVFILPINSALNPILYTLTRPKEIHRFWMYNRKSMDSKGGKTAPSF 717
 |||||
 QY 683 IWEEMWPLQEMPELMPDLFTYPCESLSISQSTRINSYS 722
 |||||
 DB 718 IWEEMWPLQEMPELMPDLFTYPCESLSISQSTRINSYS 757
 |||||
 RESULT 2
 Q912Z5 PRELIMINARY; PRT; 737 AA.
 AC 0912Z5;
 DT 01-DEC-2001 (Tremblurel, 19, Created)
 DT 01-DEC-2001 (Tremblurel, 19, last sequence update)
 DT 01-DEC-2001 (Tremblurel, 19, last annotation update)
 G PROTEIN COUPLED RECEPTOR AFFECTING TESTICULAR DESCENT.
 GN GREAT.
 OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6;
 RX MEDLINE=21250990; PubMed=11353515;
 RA Overbeek P.A., Gorlov I.P., Sutherland R.W., Houston J.B.,
 RA Harrison W.R., Boettger-Tong H.L., Bishop C.E., Agoulrik A.I.,
 RA "A transgenic insertion causing cryptorchidism in mice";
 RL Genesis 30:26-35(2001).
 DR EMBL: AF346501; AAL08943.1; .
 KM Receptor.
 SQ SEQUENCE 737 AA; 82943 MW; AF60F635EALACE49 CRC64;

Query Match 49.3%; Score 1891.5; DB 11; Length 737;
 Best Local Similarity 52.0%; Pred. No. 1.2e-133;
 Matches 370; Conservative 118; Mismatches 181; Indels 43; Gaps 7;

QY 27 CSLGTPCGNTKCLPQLLHCNGVDCCGADDEDNC-----VVVL-- 66
 |||||
 DB 28 CPKGYPCGNTKCLPQLLHCNGVDCCGADDEDNCGDTSGWTTIFGYHGVNKKVTLQ 87
 |||||
 QY 67 -----CQMSLPGLELDMKP-PFTSVPSVSSNVTAMSIQWNLIRKLPDCKRKNY 115
 |||||
 DB 88 ECPLSYPOHCYCRE--NELECVKADLAVPVSSNVTLLSLKKKHIRLPKVSRT 144
 |||||
 QY 116 DLQKLDLQNNKITSISYAFRGINSLTLYLSNRTPELPGVEEDLHREMLITEDNHL 175
 :
 DB 145 ELKRYLQNNCTHISRRAPFLGHLNQLIYLSHNCITSRPGIFKDLQHLAWLITDNDI 204
 |||||
 QY 176 SRISPPFVGLNSLILLVLMNNVLTFLPDKPLCQHPRLHMLDEGNHINHLNLTFTISC 235
 |||||
 DB 205 TRISQKSPFQNLNLFPLPWGNLEALPE-TLCAQPMQWLVLANNGIKYITNSFTLC 263
 |||||
 QY 236 SNLTFLVMRKNKINHLENTFAPLQKIDELDGSNKIENLPILFDKLKLSQNLNSYN 295
 |||||
 DB 264 DSITVFLPRLNQGIFPERKFTSSLNKNGELDSSNITKPLVHFLSDHLQKLNLSNP 323
 |||||
 QY 296 IQRQANQFYLKKSLSLEGIEISNIOORMRPLMNSHIEFKKFOYCYAPHYRSC 355
 |||||
 DB 324 LLYVHKQFQSLKQLOSLDERIEIPNISGFMQPMKNSHILKFTFYCYVPHNICM 383
 |||||
 QY 356 PNPDGSIENLNASIQRFVWVSAVYCFGNIFVYCMRPYIRSEKRLYAMSIIISLCA 415
 |||||
 DB 384 PSTDGISSSEDLNGLRVSANVAFITCVGNFLYAVSLKAEYTTAMSKILICA 443
 |||||
 QY 416 DCLMGITLFIYIGGFDLKFGEYKHAQLMMESEHCOLVGLSLSTFVSALLFTFLTEK 475
 |||||
 DB 444 DCLMGITLFIYIGGFDLKFGEYKHAQLMMESEHCOLVGLSLSTFVSALLFTFLTEK 503
 |||||
 QY 476 YCIVPERCVRGKCTIVLLIWIITGRTVAFIPLSNKEFFKNYGTNGVCFPLHSE 535
 |||||
 DB 504 FLIVYVPEFSSMLRGKQTAVALASIWVVGFLIAPPTRDYCGNFEGKKGVCPLHYDQ 563
 |||||
 QY 536 TESTGNOIYVAIFLGINLAFFIIVSYGSMFYVHQAATATEIRNOYKKEMLAKR 595
 |||||
 DB 564 AEDPGSGYSLGIFLGNLALFLVIVSYVTMCSIKHTALQAEVASHGKEVAVANRF 623
 |||||
 QY 596 FFIVFTDALCWPILFVVKFSLQVEIIGTTSWVVFILPINSALNPILYTLTRPKE 655
 |||||
 DB 624 FFIVFTDALCWPILFVVKFSLQVEIIGTTSWVVFILPINSALNPILYTLTRPKE 683
 |||||
 QY 656 MIHREYTYNRKSMDSKGGKTAPSF IWEEMWPLQ-----EMPELMPKP 700
 |||||
 DB 684 LKQDLHKHR-RKPIFKVKKRSLASIVWTDESSLKLVLSKIALGDSIMKP 734
 |||||
 RESULT 3
 Q9VBPO PRELIMINARY; PRT; 359 AA.
 AC Q9VBPO;
 Q9VBPO;

DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update).
DE 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE CS5042 PROTEIN.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY.
RC MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celiker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Adell J.F., Agbayani A., An H.-J., Andrews-Plannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brostein P., Brothier P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fierlich K., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jallat M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lascko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Paclob J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stepleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weisslock G.M., Weissensbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC EMBL: AE003753; AAF56490.1; -
DR FlyBase; FBgn0039354; CG5042.
DR InterPro; IPR002106; AA_trna_ligase-II.
DR InterPro; IPR000276; GPCR_Rhodopsn.
DR Pfam; PF00001; 7tm_1; 1.
DR PRINTS; PRO0237; GPCR_Rhodopsn.
DR PROSITE; PS00339; AA-trna_ligase-II_2; UNKNOWN_1.
DR PROSITE; PS00237; G-PROTEIN_RECPT_FL_1; 1.
DR PROSITE; PS0262; G-PROTEIN_RECPT_FL_2; 1.
KW G-protein coupled receptor; glycoprotein; transmembrane.
SQ SEQUENCE 359 AA; 40880 MW; 71913BEBD7C9E739 CRC64;

Query Match 15.6%; Score 599.5; DB 5; Length 359;
Best Local Similarity 39.9%; Pred. No. 4e-37;
Matches 132; Conservative 60; Mismatches 124; Indels 15; Gaps 7;

OY 349 PVRMCKPSTGDSVSSFDLLSKPVLRYSAWMAATLTIAGNVLVAGRPYRDENVAVTMV 408

Db 3 PVRMCKPSTGDSVSSFDLLSKPVLRYSAWMAATLTIAGNVLVAGRPYRDENVAVTMV 62
OY 409 IISLCADCLMGITLVYIGGFDLKFGRGYNNKHAQIMESTHCQIVGSAIISTEVSULL 468
Db 63 IRLNLADMLMGFLVTVIGVODYRKYRVNLDWITSMQCTLTGLTAVSSSEVSMIL 122
OY 469 TELTLEKICIVYPRCPKCRIT-TVLIVITGPIVAFIPL-----SNKEPFKYY 522
Db 123 AFMSLERLTLADPPRGHRSIGNRYMMLALCMTITGVLAVAPVLLMRTSTLPYGSYS 182
OY 523 GTNGVCEPLHSEDTEISGAQIVSAIFGLINIAAFIIVFSYGSMPYSVHOSATATEIR 582
Db 183 GT---CFPLHHEAFPMG-WLYSAFVFLVGNLLVLMATLVALLISWRTS-SATPL- 236
OY 583 NOVKKEMLARFFPIYTDLCWPIPVVAFSLQVEIGTITSWVITIPINSALN 642
Db 237 --TLIDCEFAVRFFPIYTDLCWPIPVVAFSLQVEIGTITSWVITIPINSALN 294
OY 643 PILVLTTRPFKEMTH-REWYNYRORXKMSDS 672
Db 295 PILVLTTRPFKTRNQLFLGKMKKITSRKRAEA 325

RESULT 4
ID 095Y16 PRELIMINARY; PRT: 1012 AA.
AC 095Y16;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE GLYCOPROTEIN HORMONE RECEPTOR.
GN ACPHR.
OS Asterina pectinifera (Starfish).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Asterozoa;
OC Asteroidea; Valvatacea; Valvatida; Asterinidae; Asterina.
OX NCBI_TaxID=7594;
RN [1]
RP SEQUENCE FROM N.A.
RA Mita M., Hirai T., Oba Y., Yoshikuni M., Nagahama Y.;
RT "cDNA cloning and functional analysis of a novel member of the
RT glycoprotein hormone receptor family from a starfish Asterina
RT pectinifera."
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB061862; BAB68209.1; -
KW Receptor.
SQ SEQUENCE 1012 AA; 112623 MW; 52A70E7A88C46E0A CRC64;

Query Match 15.0%; Score 576; DB 5; Length 1012;
Best Local Similarity 25.9%; Pred. NO. 7.7e-35;
Matches 193; Conservative 102; Mismatches 262; Indels 188; Gaps 21;

OY 83 FTSVSVS-----SNVAMSLOMNLIRKLPDCEKYNHDLQKLDLQNNKITSISIVAFR 136
Db 127 FQGYPRKAFRNDDLANLRKHLDSMWIREVPADAMNLTALHHLWLDHNSQSEVFTALH 186
OY 137 GINSITKLYSHNRITPKGVF-EDLHRLEWLIIEDNHLSTISPPPYGLNSLLVLVLM 195
Db 187 HLSNRIILHENSIPVVDHAFENSHLIE-LILRHKKITHLSAARVAGLPLNMLEFL 245
OY 196 NNVLRL-----PD-----KPLCQ 209
Db 246 GNSITSIHNTAFRNLPALRNVLIVLKVLSVPPDLGTTSLEHLIERCSLRAITANFCD 305
OY 210 HMPRLHWLDLEGNHINLNLITFISCSNLTLYVMKRNKINHLNENTFAPLOKIDELDGS 269
Db 306 NMTGLTSLNLHNLLEGPSLS-KCSSLKVLHCTNKLKTLSEGGPFSGLHLYDQLLE 363
OY 270 NKEIENPLIFKDLKELSQLNSYNPIQIQANQDYLVKLSLE-----GIE 319
Db 364 NDISYIPADAFOSLSHLTLTSLSNNTIREIDSQAFPCISLOYLDLSNNSFVPLTAGIQ 423
OY 320 I-----SNIQQRNFRPLMNLSHI-----YFKFQYQY 347

```

Db 424 MLKIRTYDNEQLEDFPPSELPSTETATAY--PYHCEYTELEAEYKSLADRNISE 481
OY 348 -----APHVSCKPN 357
Db 482 TTYWASGVDPDYNNMTFDINSESWIDSIFGFSLSIGSPITYLSGNSRLVYHNHNSCRK 541
OY 358 TDGSISSLENLAIIOREVVWVYSAVTCGNIYICMRPYINSEKLYAMSIISLCCADC 417
Db 542 PGPFMPCMDLFGSWPLRIGWLVFLAIIIGNAIVFVIVSHTKMDVPREFLIGNLAFADF 601
OY 418 IMGITLVFGGDLKFRGYKNKHAOLMSESTHCQLVGSALISTEVSVLLFLFLEKYI 477
Db 602 FLGYVLGFLAGVDTSTLGVFRKGARWQLSAGCRLAGFLAVSSEFSITLSTYTLERY 661
OY 478 CIYVPRCVRPGKC-RTITVLIIWITGFIYAFIPLSNKEFFKNYGTNGVCFPLHSED 536
Db 662 AIKHAHLERKMKLPRAITVMOGWIIFSVTAVPLVNV---VSHHRAVPCLPF---DV 714
OY 537 ESIGAOIYVAIFLGINLAFTIIVFSGSMFYSHQS-AITATEIRNOYKKEMLAKRF 595
Db 715 DYTVAKVY-VGSILLIINLAIYIMACVASYIALOGSHAMNCNDR-----VARRK 765
OY 596 FTIVFTDALCWIPI-----FVVKFSLQVEIGTITSWVIFILPINSALNPLYT 647
Db 766 SLIVFTDFACWADIAPFSITLAFGLRLISLDGAKV-----LTIIVPLNSCANPFIYT 818
OY 648 LITRPERK---EMIRFPWYN--YRQR 667
Db 819 ILTKQFKKCKOCTIMRSLNSNVFRQR 843

```

```

RESULT 5
O95Y17 PRELIMINARY; PRT; 1280 AA.
ID 095Y17:
AC 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE GLYCOPROTEIN HORMONE RECEPTOR.
GN ACPGR.
OS Asterina pectinifera (Starfish).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Asterozoa;
OC Asteroidea; Valvatacea; Valvatida; Asterinidae; Asterina.
OX NCBI_Taxid=7594;
RN [1]
RP SEQUENCE FROM N.A.
RA Mita M., Hirai T., Oba Y., Yoshikuni M., Nagahama Y.;
RT "cDNA cloning and functional analysis of a novel member of the
RT glycoprotein hormone receptor family from a starfish Asterina
RT pectinifera."
RT Submitted (May-2001) to the EMBL/GenBank/DBJ databases.
RL EMBL; AB061861; BAB68208.1; -.
KM Receptor.
SQ SEQUENCE 1280 AA; 141700 MW; 0AB0ECC0DD880BA CRC64;

```

```

Query Match 15.0%; Score 576; DB 5; Length 1280;
Best local Similarity 25.9%; Pred. No. 1e-34;
Matches 193; Conservative 102; Mismatches 262; Indels 188; Gaps 21;
OY 83 FTGVSPTS-----SNVTAMSLQNLIRKLPDCEFKNYHDLQKIDLNKNTISITAFR 136
Db 127 FOQVPRKAFRNDLANLRKILDSNMWIREVPADAFNMLTALHINLNDHNLSEVPTAALH 186
OY 137 GLASITFLYLSHRITLTKGVF-EDHRLKEMLIEDNHSRISPPFYGINSILILVLM 195
Db 187 HLSNRLILHEHNSIPVDPHAFENSHLE-LILRNKITHSAHAFAGLIMLLEFL 245
OY 196 NVNLTFL-----PD-----KPLCQ 209
Db 246 GNSITSTIAHTAFRLPALRLNVLLEVKNLSVFPDLTGTSTLSLHGIRCSLRALPAPFCD 305

```

```

OY 210 HMRPLHMLDLEGNHINLNRNLFICSNLTVLMRKKNINHNENTFAFLQKLDELIGS 269
Db 306 NMTGLTSLNHNHNLIEGLPSLS--KCSSLKVHLGNTKNTLSLEGQPGFGLHLYQLLE 363
OY 270 NKIENLPLPIFDKLEQLSOLNSYNPIQIOANOFDYLKLSLE-----GIE 319
Db 364 NDISYIPDAFQSLSHDPTLSNNMTIREIDQAFAPCTSLQYLLDLSNNSFPVPLTAGIO 423
OY 320 I-----SNIOQRMRPLMNSHT-----YFKFQYQGY----- 347
Db 424 MLKIRTYDNEQLEDFPPSELPSTETATAY--PYHCEYTELEAEYKSLADRNISE 481
OY 348 -----APHVSCKPN 357
Db 482 TTYWASGVDPDYNNMTFDINSESWIDSIFGFSLSIGSPITYLSGNSRLVYHNHNSCRK 541
OY 358 TDGSISSLENLAIIOREVVWVYSAVTCGNIYICMRPYINSEKLYAMSIISLCCADC 417
Db 542 PGPFMPCMDLFGSWPLRIGWLVFLAIIIGNAIVFVIVSHTKMDVPREFLIGNLAFADF 601
OY 418 IMGITLVFGGDLKFRGYKNKHAOLMSESTHCQLVGSALISTEVSVLLFLFLEKYI 477
Db 602 FLGYVLGFLAGVDTSTLGVFRKGARWQLSAGCRLAGFLAVSSEFSITLSTYTLERY 661
OY 478 CIYVPRCVRPGKC-RTITVLIIWITGFIYAFIPLSNKEFFKNYGTNGVCFPLHSED 536
Db 662 AIKHAHLERKMKLPRAITVMOGWIIFSVTAVPLVNV---VSHHRAVPCLPF---DV 714
OY 537 ESIGAOIYVAIFLGINLAFTIIVFSGSMFYSHQS-AITATEIRNOYKKEMLAKRF 595
Db 715 DYTVAKVY-VGSILLIINLAIYIMACVASYIALOGSHAMNCNDR-----VARRK 765
OY 596 FTIVFTDALCWIPI-----FVVKFSLQVEIGTITSWVIFILPINSALNPLYT 647
Db 766 SLIVFTDFACWADIAPFSITLAFGLRLISLDGAKV-----LTIIVPLNSCANPFIYT 818
OY 648 LITRPERK---EMIRFPWYN--YRQR 667
Db 819 ILTKQFKKCKOCTIMRSLNSNVFRQR 843

```

```

RESULT 6
O9BN18 PRELIMINARY; PRT; 1050 AA.
ID 09BN18:
AC 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE LEUCINE-RICH REPEAT-CONTAINING G PROTEIN-COUPLED RECEPTOR 2.
GN RK.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_Taxid=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Nishl S., Hsu S.Y., Zell K., Hsueh A.J.;
RT "Characterization of two fly LGR (leucine-rich repeat-containing G
RT protein-coupled receptor) proteins homologous to vertebrate
RT glycoprotein hormone receptors: constitutively activation of wild type
RT fly LGR1 but not LGR2 in transfected mammalian cells."
RT Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RL EMBL; AF274591; AAK00808.1; -.
Db InterPro: IPR00276; GPCR_Rhodpsn.
DR InterPro: IPR001611; LRR.
DR InterPro: IPR003592; LRR.
DR InterPro: IPR003591; LRR_tyr.
DR Pfam: PF00001; 7tm_1; 1.
DR Pfam: PF00560; LRR; 10.
DR PRINTS: PR00237; GPCRHHODPSN.
DR SMART: SM00370; LRR; 8.
DR SMART: SM00369; LRR_tyr; 10.

```


QY 574 SATTATEIRNOYKEMILAKREFEIVFDALCMPIFVVKFSLLOVELPGTTSWVIP 633
DB 569 PSSAAAHADTRV-----AQRMAVLFTPTDTCNAPISFPAISAAKLPL-ITYSDSKLL 621
QY 634 IL--PINSALNPLTYLTRPFK 654
DB 622 VLEFPINCSNPFLYAFETRNFR 644
RESULT 8
Q9NRK06 PRELIMINARY: PRT: 1300 AA.
ID Q9NRK06 Q9VJ03:
AC Q9NRK06: Q9VJ03:
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE HYPOTHEICAL.144.0 KDA PROTEIN (RK GENE PRODUCT).
GN RK OR BG:DS00180.13 OR CG8930.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
NCBI_TaxId=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Y, AND CN BW SP;
RX MEDLINE=9403001: PubMed-10471707;
RA Ashburner M., Mitra S., Roote J., Lewis S.E., Blazej R., Davis T.,
RA Doyle C., Galle R., George R., Harris N., Hartzell G., Harvey D.,
RA Hong L., Houston K., Hoskins R., Johnson G., Martin C., Moshrefi A.,
RA Palazolo M., Reese M.G., Spradling A., Tsang G., Wan K., Whitehead K.,
RA Celinker S., Rubin G.M.;
RT "An exploration of the sequence of a 2.9-Mb region of the genome of
RT Drosophila melanogaster: the Adh region.";
RL Genetics 153:179-219(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Y, AND CN BW SP;
RX MEDLINE=20196006: PubMed-10731132;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brannon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miles G.L.G.,
RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borstova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Burks K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durkin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hoslin D., Houston K.A., Howland T.J., Wei M.-H., Idegami C.,
RA Jalili M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kuip D., Lai Z.,

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Moberly C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Munzy D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusslein D.R., Puri V., Reese M.G.,
RA Palazolo M., Pitman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhao M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL: AE003408; AAF4846.1; -;
DR EMBL: AE003642; AAF5367.2; -;
DR FLYBASE: FBgn0003255; rK.
DR InterPro: IPR000276; GPCR_Rhodopsn.
DR InterPro: IPR003592; LRR_out.
DR InterPro: IPR003591; LRR_typ.
DR Pfam: PR00001: 7tm.1: 1.
DR PRINTS: PR00237; GPCR_RHODOPSIN.
DR SMART: SM00370; LRR; 1.
DR SMART: SM00369; LRR_TYP; 4.
DR PROSITE: PS00262; G_PROTEIN_RECEP_F1_2; 1.
KW Hypothetical protein.
SO SEQUENCE 1300 AA; 144031 MW; B4B9E39F942FA0B3 CRC64;

Query Match 12.8%; Score 492.5; DB 5; Length 1300;

Best Local Similarity 22.6%; Pred. No. 1.9e-28;

Matches 187; Conservative 110; Mismatches 292; Indels 237; Gaps 21;

QY 45 LMCNGYDCGNQADENDNCVYL-----CQMSLPGLE 76
DB 163 LSCRGIGIILAVPNLPENVVLTLSNDSIINMDFYGLAKLRSLQNGCISLPPOS 222
QY 77 LDMKRFSTSPVSSN-----VTAMSLQNLIRKLPDPCKNVHDKLD 121
DB 223 FQGLAULTSL-QUNGALVSLDGDGHLQKLTIRKGNLFRTIPNALAGRTLEALL 261
QY 122 LQNNKRTISIVFRLNSLTLYLSHNRITFLKPGV----FEDL-----HRLW----- 167
DB 282 LKRNQMKISAGALKNLTLAKVLELDNLISSLPGLSKISQGLSTISNRRLWINDTE 341
QY 168 -----LITEDNHSRISPFYGLNSLILVLMN-NVITRLPD----- 204
DB 342 LPRSMQMLMRANPLDLSISPGARSGMSKLRKLTISDVRTLSPPELEACHALEILKLDRA 401
QY 205 -----KPLCOHNPRLAHMLDLEGNHILNRLNLFISCSNLTIVYMKRNKINHLENTFAP 258
DB 402 GIOEVPANLCROTFRKSLDLSQNIQKPGFNGIKQLNDLILSYNRKALPQAFQ 461
QY 259 LQKLDEL-----DLGSNKTEMLPPLFKDELKELSQLNLSYNPQIKQANQPDYLV 308
DB 462 IPRQLQLVYSIIRIHSDELGNSIYIKRAFGFPALEDNLGNINIPFLPESGLRALL 521
QY 309 KKLKSLSEGI-----EISNIQQRMRP-- 330
DB 522 HKTFFNPKLRPPPPDFPRPRTIQLTILSYAHCCAFPLVAMSSQKTSQVQDAVLPSPD 581
QY 331 -----LMLN-----SHI-- 337
DB 582 AEFDMTLNNSNMNIPOMHNLISKQLGASMDPWEALNFNEBQLOTQGGQATATSYME 641
QY 338 YPKFQYCG-----YAPHVRSCKPNTDGSSLENTLASTIQHVFVW 378
DB 642 YFEHHVSGPAICGCTGLFSCMSTEDPQGSVQCLPAPGPPLPADLPDMWTTLACGW 701
QY 379 VVSAYVTCFQGNIFVCMRPYIRSENKLYAMSIISLCCADCIMGIYLFVIGFDLKRGEYN 438

```

Db 702 VVELLSLNGCYVYVLLCSRSKMDVPRFLVCNLAADPFMGITGLALVDAATLGEFR 761
QY 439 KHAQUMESTHCQLVSLALSTEVSVLLTFLTEKYICYVPERCVRGKCTITVLI 498
Db 762 MFAIPQMVSVALCOLSGELAVLSSELVYTLAVITLERVVAITHAHLKRLSLKQACYIM 821
QY 499 LI-WINGFIYAFLPLSKKEFFKNYNGVCFPLHSDTESIGQIYSVAIFLGINLAIF 557
Db 822 SVGVFVALMALMPVGVSDYRKF---AVCLPE--ETTTGPASLTYYVISLMF--INCGAF 874
QY 558 IIVFYSGSMFYSHOSAITATEIRNOVKEMILAKREFEVLFDALCWTIP-----I 609
Db 875 LITLGCYLCIKYMAIRGSOANT-----NDSRIAKMALIVFDLFCWSPFIASFITAI 927
QY 610 FVVKFSLQLVEIPGTTTSMVVFILPINSALNPILYTLTPRKE 655
Db 928 FGLQISLSEQAKI-----FTYFVFLPNSCNPFLYAIMTKQKK 966

RESULT 9
Q9Z1P4 PRELIMINARY; PRT; 907 AA.
AC Q9Z1P4;
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DE 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE ORPHAN G PROTEIN-COUPLED RECEPTOR FEK.
GN GPR49.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RS SEQUENCE FROM N.A.
RA MEDLINE=99121227; PubMed=9920770;
RA Hermy G., Methner A., Schaller H.C., Hermans-Borgmeyer I.;
RT "Identification of a novel seven-transmembrane receptor with homology
RT to glycoprotein receptors and its expression in the adult and
RT developing mouse."
RL Biochem. Biophys. Res. Commun. 254:273-279(1999).
DR HSSP: P23945; 1XUN.
DR MGD: MGI:1341817; Gpr49.
DR InterPro: IPR000276; GPCR_Rhodopsn.
DR InterPro: IPR001611; LRR.
DR InterPro: IPR000372; LRR_Nterm.
DR InterPro: IPR003592; LRR_out.
DR InterPro: IPR003591; LRR_tyr.
DR Pfam: PF00001; 7tm.1; 1.
DR Pfam: PF00560; LRR.15.
DR Pfam: PF01462; LRRNT.1.
DR PRINTS: PR00237; GPCR_Rhodopsn.
DR PRINTS: PR00019; LEURICHROPT.
DR SMART: SM00370; LRR.3.
DR SMART: SM00013; LRRNT.1.
DR SMART: SM00369; LRR_tyr.8.
DR PROSITE: PS0262; G_PROTEIN_RECEPTOR_FL_2.1.
KW Receptor.
SQ SEQUENCE 907 AA; 99681 MW; 5531676C60AAE253 CRC64;

Query Match 12.8%; Score 490; DB 11; Length 907;
Best Local Similarity 24.6%; Pred. No. 1.9e-28;
Matches 182; Conservative 110; Mismatches 263; Indels 184; Gaps 24;

```

```

QY 188 SLILVLMNVLRLP-----DKPL-----CQMPRLHMLDLS-NH 223
Db 258 NLKEGFSNNKIRISPERAFVGNPSLTIHFYDNPIDFVGVSFAOHLPBELRTLTLANASH 317
QY 224 IHNLRNLTFSICSNLTVLVARKKNKINHLNMENTFAPLOKLDLSDGSKINLPLIFKDL 283
Db 318 ITPFPNLT--GTATLESLLITGAKISLPAVCDQPLNLOYLDLSTYLLDLSL--SGC 373
QY 284 KELSQNLSTNPNLOKIQANOFDYLVKLKSLSLGEIETSNIOQRMFLPM----- 333
Db 374 OKLOKIDLRNEIETIEIGSFQOLFNRSLNLMANKIATIHHPANFSPLPLIKDLSSNL 433
QY 334 -----LSHTY-----KKFYCGY----- 347
Db 434 LSSEFVYGLHGLTKLGNRALOSLIPSANFPELKIITEMPSAVQCCAFGCENYKISN 493
QY 348 -----APHVRSCKPNTDGISLEN 366
Db 494 QMKNKDGNSVDLHKKQAGLFQVODDERDLEDPLDPEEDLNALHSVCCSPSPGPFKCEH 553
QY 367 LLSAIIQVFEVWVSAVTCGNIFFV---ICMPYIRSENKLYAMSIISLCCADCLMGYL 423
Db 554 LFGSMILRIGWMTAVTLTSONMLVALTVFRPLYSISIKL-LIGVIAV--VDIIMGVSS 610
QY 424 FVIGGFPLKRGERYNKAQUMESTHCQLVGSALSTEVSVLLTFLTEKYICYYPF 483
Db 611 AVLAAYDAFTFGFRAQHGWAMWEDIGCOIVGFLSIFASESIFLTLAALRGVSV--- 666
QY 484 RCVAPGKC-----RTTVL-ILVITGFIYAFIPISNKEFFKNYGTNVCF 529
Db 667 -----KCSKFEVKALPFSIRALVILCVLALT---IATIPPLG-----GSKYNASPLCL 713
QY 530 PLHSEDTESIGAQIYSVAIFLGINLAIFIIVESGSMFYSHOSAITATEIRNOVKEM 589
Db 714 PLRGPESITG--YNAVALV-LNSLCFLIMTIYATKLYCSLEK-----ELEVLMDCSM 764
QY 590 ILAKREFEYFTDALCWTIPFVVKFSLQVE-IPGITSVMVFIPLINSALNPITLYL 648
Db 765 V--KHIALPLFANCLICPAVAFSSSLNLTFTSPVINFILVILVPLSCINPLLYIV 822
QY 649 TTRPFKEMI-----HRFM 661
Db 823 FNPFRKEDMSLGKHTRFW 841

RESULT 10
Q9NDI1 PRELIMINARY; PRT; 1360 AA.
AC Q9NDI1;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE GLYCOPROTEIN HORMONE RECEPTOR II.
GN RK OR BG:DS00180.13 OR CG8930.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydriodea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RS SEQUENCE FROM N.A.
RA STRAIR-CANTON S.; TISSUE=WHOLE ANIMAL;
RA MEDLINE=20359836; PubMed=10899142;
RA Eiksen K.K., Hauser F., Schiott M., Pedersen K.-M., Seendergaard L.,
RA Grimmelikhuijzen C.J.P.;
RT "Molecular Cloning, Genomic Organization, Developmental Regulation,
RT and a Knock-Out Mutant of a Novel Leu-Rich Repeats-containing G
RT protein-coupled Receptor (DLCR-2) from Drosophila melanogaster."
RL Genome Res. 10:924-938(2000).
DR EMBL: AF142343; AAF66608.1; -.
DR HSSP: Q57815; 1D3Y.
DR FlyBase: FBgn0003255; rk.

```


QY	410	ISLCAACGILGIVFYVGGEDLKRGRGYNHAOLAMSTHCOLVGLAIIESTEVSLTLT	4659
Db	381	CHSLFADLCMGIVLYIYATAYDVRKGLYYNHAISMWQAGCDIAGFTTVASELSMTLT	4400
QY	470	FLELEKYICIVPEPCVRPKCR-TITVLLIWTGFIYAFIPLSNKEFFKNYGTNGVC	5288
Db	441	AITDERHTTTHALRLDRKRLRLHACAVMTAGMASCTALLLPTVG---VSSYSKVIC	4566
QY	529	FPLHSEPTESGAIOYSVALFGLGNLAFLIIIVESYSGMSYVHQSAITATETIRNOYKKE	5688
Db	497	LPM---DVESLPSGVF-VMLLELLNVAFGLVCVCYIISLVSNSS-----SPASAE	5466
QY	589	MIARREFEYITATLALCMIDFVYVKKFSLILOVEIPGTITSMVYFII--PINSALNPLY	6466
Db	547	TCMAQRMAIILITDFELCMARISPFALSAAIKLPII-IVDSKILLIYLFYINSCANFLY	6005
QY	647	TLTTRPKEMIHFRWYINRQKBS-DSKQCKTAPSR-----IWEKMP	689
Db	606	GLCIRTR---RDGLIARVGLFTTKRAVYRTRESVSQAOAMIOISP	650

	RESULT	12
ID	09DGC5	PRELIMINARY; PRT; 693 AA.
AC	09DGC5;	
DT	01-MAR-2001 (TrEMBLrel. 16, Created)	
DT	01-MAR-2001 (TrEMBLrel. 16, Last sequence update)	
DE	01-DIC-2001 (TrEMBLrel. 19, Last annotation update)	
GN	GONADOTROPIN RECEPTOR II.	
OS	TGTH-R11.	
OC	Oreochromis niloticus (Nile tilapia) (Tilapia nilotica)	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostei; Neopterygii; Neoplethyrini; Teleostei; Euteleostei; Perciformes; Acanthomorphi; Acanthopterygii; Percomorpha; Perciformes; Acanthidae; Oreochromis.	
OX	NCBI_TaxID=8128;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RA	Oda Y., Hirai T., Yoshitara Y., Yao Z., Nagahama Y.;	
RT	"Tilapia gonadotropin receptor II.";	
RL	Submitted (APR-2000) to the EMBL/GenBank/DBJ databases	
DR	EMBL; AB041763; BAB16107.1; -	
DR	HSSP; P22888; ILTU.	
DR	InterPro; IPR000276; GPCR_Rhodopsn.	
DR	Interpro; IPR001611; LRR.	
DR	Pfam; PF00001; 7tm_1; 1.	
DR	Pfam; PF00560; LRR; 2.	
DR	PRINTS; PRO0237; GPCRHODOPSN.	
DR	PROSITE; PS00237; G_PROTEIN_RECIP_FL1; UNKNOWN_1.	
DR	PROSITE; PS0262; G_PROTEIN_RECIP_FL2; 1.	
KW	SIGNAL.	
QO	SCQUENCE. 693 AA; 77856 MW; C3BBEFD8ECFC8988 CRC64;	

Query Match	12.5%	Score 481	DB 13	Length 693
Best Local Similarity	25.4%	Pred. NO. 6.6e-28		
Matches 156; Conservative	118;	Mismatches 235;	Indels 104;	Gaps 24;

[illegible]

OY	320	- - - - - : I N O O M E R P L A N - - - - - L S H Y K F K E F U - - - - - C G Y
Dd	280	S H C C A F T W R K R O E S L A K N U T K - F D D L M N T E I D P P A D O T S L I N D I N F O Y R D L E R D C F S
OY	348	A P H V R S C P M T D G I S S E N L A S I J O R E V F W V V S A V C F G N I E V I C M R P R I R S E N K U Y A M
Dd	338	N P E F K - C S P R P D A N P C E D L L G F S F L X C L M I I M V E F A V A G N L A V I - L L I H N K L Y S
OY	408	S I T - S L C C A D C L M G I L F V I Y G G F D L K R B E Y K N K A O L M E S T H C O L V G S L A I L S T E V Y
Dd	395	R E L M C N I A F D L C G L X L I I L I A F E D Y S H H E Y N H A D M Q T G C G C I A G L T V F S S E L S V
OY	466	L L L F L L E K Y I C I V F R C V R P K C K T I V L I - I W T F I Y A F P L S K E F F K N Y G T
Dd	455	Y T L I V I S L E R H T T L N A M H V K R L R M H N H V A M V G M A F S L L A L L P L V G - - - - - V S S Y S K
OY	525	N G V C F P L S E D T E S I G A O I Y S V A L F I C I N T A F I I Y S G M F Y S H O S A I T A T E R N Q
Dd	511	V S I C L P M - - - D I D L G A O V Y V A N L I - L N Y A F L V O C Y C I C I T S L V S H N B H S R - R G D
OY	585	V K K E M I A K R E F F E V T D A L C W I P F V V K L S L O V E I P E T T S W V I I L - P I N S A L N
Dd	565	T K - - - - - I A K R A V A L I F T D F C M A P I S F F A I S A L R M P L I - T V S H S K I L L I F Y P I N S L C N
OY	643	P I Y L T L T R P P K E
Dd	620	P E L Y I T F T R A E R K

RESULT	13
09DGF5	
ID	09DGF5
AC	09DGF5;
DT	01-MAR-2001 (TREMBLrel. 16, Created)
DT	01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE	FOLLICLE-STIMULATING HORMONE RECEPTOR PRECURSOR.
OS	CYNOPS PYRRHOGASTER (Japanese common name).
NC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Amphibia; Batrachia; Caudata; Salamandroidae; Salamandridae; Cynops
OX	NCBI_TaxId=6330;
RP	[1]
RN	SEQUENCE FROM N.A.
RC	TISSUE=TESTIS;
RC	MEDLINE=20403884; PubMed=10944452;
RA	Nakayama Y., Yamamoto T., Oba Y., Negahama Y., Abe S.-I.;
RT	"Molecular cloning, functional characterization, and gene expression
RT	of a follicle-stimulating hormone receptor in the testis of newt
RT	Cynops pyrrhogaster.";
TL	Biochem. Biophys. Res. Commun. 275:121-128(2000).

DR	HSSP, p23945; 1XUN.
DR	InterPro; IPR000276; GPCR_Rhodpsn.
DR	InterPro; IPR001611; LRR.
DR	InterPro; IPR000372; LRR_Nterm.
DR	Pfam; PF006001; 7tm_1; 1.
DR	Pfam; PF00560; LRR; 4.
DR	PRINTS; PR00237; GPCRHDOPSN.
DR	SMART; SM00013; LRRNT; 1.
DR	PROSITE; PS00237; G_PROTEIN_RECIP_F1_1; UNKNOWN_1.
DR	PROSITE; PS0262; G_PROTEIN_RECIP_F1_2; 1.
KM	Signal; Receptor.
FT	SIGNAL
ET	CHAIN 1 17 POTENTIAL.
EQ	SEQUENCE 696 AA; 78633 MW; 1/9A6FC8B00B71E57 CRC64

Query Match	12.5%,	Score 479.5;	DB 13;	Length 656;
Best Local Similarity	23.7%;	Pred. No. 8.6e-28;		
Matches 165;	Conservative 133;	Mismatches 250;	Indels 149;	Gaps 30.
OY	32	PFGGNTTCLEPQLHCHGNAVDDCGNOADEDCNVVULCQCMSLEGLIEDMKKPTTSVP-SVS	90	

```
Db 16 FCGHPVCRCLNRYETC-----QESHVY-----QIPRDIP 44
QY 91 SNTYAMSLQNNLRKLPDPCCKNHYDLOKLDL-QNNKITSIYAFAKGLNLTLYLSH- 148
Db 45 RNSLETFVLTQVIVIPKAAFGSEDEVENIESQNDLAKTIEANVFHLLPLRLRIEKA 104
QY 149 NRITFLKPGVEDLHRLLEWLIIEDNHLRSISPPFYGUNSLILLYLNNVLTRLPDLPC 208
Db 105 NNLVYIDPDAFQNPRLKYLILSTNGIOY--PAVSKIRS-----PHSVL----- 147
QY 209 QHMRPLHMLDEGN-HIHNLRNLTFISCSNLVLYV-MRKKINHLENFPAPOKLDLDEL 266
Db 148 -----LIVQDINIRHIGKNSFAGLSSESTIRLNKNGIEIQHNAENGTH-LNELN 198
QY 267 LGSN-KIENLPLIFKD-----LKEISQNLNLYNP-----IQIQANQDLYLVKLSLS 314
Db 199 LSDNQRLKLPDOVFQATGPVILDISRTRIHLELPNNGIENIKKFRARFNYLKLPLPE 258
QY 315 -LEGIE-----ISNIOQM--FRPLMN-----L 334
Db 259 KPAELLIANLTYPSHCCAFAANRERKSEMPICNKSFGKHSACKPEDKNLRRESNEDYL 318
QY 335 SHIYFK-----KFOY---CGYADHVRSCKPNIDGISSLENLASSIQRVFVWVSA 382
Db 319 SSYGFSTSLVENGDENFYDYLCEV--HDVYICPEKPDAPNCEDEDIMQDNTLRVLIWLSI 377
QY 383 VTCEGNLFVCMRPYIASENKKIYAMSTII--SLCCADCLMGITLFLVIGDPOLKPFGEVKNH 440
Db 378 LAITGNITVAVI--LISSQKFTVPFRLMCLNAPADLCMGITLYLLILASVDTKRSQYVNH 435
QY 441 AQLMESTHCQLVSLAILSTEVSVLLTFLETKYICYVPCFVCPGKCRITVYILI 500
Db 436 AIDMQTSGCAAGFEVFPVASELSVTLVITLERMHTITTYAMQLDKVRFRHATAMV 495
QY 501 -WITGFVAITPLSNKEFFKNGYGTNGVCPPLHSEDTEISGAQIYVAIFLGINLAFTI 559
Db 496 GWTFAPFVALPLI---FGVSSYTKVASICLPM---DIESRISOSIYFI-LVLWVCAFLI 547
QY 560 IYVSYGSMFYSHQSAITTAITEIRNQVKEMILAKRFFIYFDALCMPIFVYVAFELLO 619
Db 548 ICACYIGIYLVTRNPNVSS-----NSDRIKAKRMALIFTDPLCMAPISFAISLSLK 601
QY 620 VEIPGTTTWNVIFIL--PINSALNPLIYLTTRPFK 654
Db 602 IPLI-TVSKSKILLVLFYPINSCANPFLVAFIFKTFR 637

RESULT 14
Q9PW16 PRELIMINARY; PRT; 662 AA.
AC 09PW16;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, last sequence update)
DE 01-JUN-2001 (Tremblrel. 17, last annotation update)
DE FOLLICLE-STIMULATING HORMONE RECEPTOR PRECURSOR.
GN FSH-R.
OS Clarias gariepinus (Sharpooth catfish) (African catfish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC Siluriformes; Clariidae; Clarias.
OX NCBI_TaxID=13013;
RN [1]
RN SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RA Bogerd J., Andersson E., Blomenrohr M., Tensen C.P., Graneman J.C.,
RA Schultz R.W., Geos H.J.;
RT Cloning and functional characterization of a testicular follicle-
RT stimulating hormone receptor of the African catfish."
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; A0102647; CAB51907.1; -.
DR HSSP; P23945; 1XUN.
DR InterPro; IPR000276; GPCR_Rhodopsn.
```

```
DR InterPro; IPR001611; LRR.
DR Pfam; PF000001; 7tm_1; 1.
DR Pfam; PF00560; LRR; 4.
DR PRINTS; PR00237; GPCRHHODPSN.
DR PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; UNKNOWN_1.
DR PROSITE; PS00262; G_PROTEIN_RECEP_F1_2; 1.
KW Signal; Receptor.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 662 FOLLICLE-STIMULATING HORMONE RECEPTOR.
SQ SEQUENCE 662 AA; 74891 MM; 5CB49EDDBD1DAF5 CRC64;

Query Match 12.5%; Score 478; DB 13; Length 662;
Best Local Similarity 25.4%; Pred. No. 1,le-27;
Matches 157; Conservative 94; Mismatches 257; Indels 110; Gaps 21;

QY 92 NVTYAMSLQNNLRKLPDPCCKNHYDLOKLDL-QNNKITSIYAFAKGLNLTLYLSHNR 150
Db 49 NTVYVEIKLQIIFPRAMSSLDLKRIMVSENGALQRIEAVAFANLITLEEI----- 102
QY 151 ITFLKPGVEDLHRLLEWLIIEDNHLRSISPPFYGUNSLILLYLNNVLTRLPDLPCQ 210
Db 103 -----TITKSKNLVSMRDTFGWLPKLTLYISNTGLVLPDPSKYQS 145
QY 211 MPRHLWLDLGN-HIHNLRNLTF--ISCSNLVLYVMKRNKINHLENFT----- 256
Db 146 AAEFLPDLDMNHIEVYHISNAFAGLISGTTTELRLTKNGITEVERNAFNGTKMEKLELM 205
QY 257 -APLOKLDL-----LDLGSNKIENLPLIFKDK----- 264
Db 206 GNOOLKRIIDHNAFLGAGPLVLDISRTAISLSPENMRILKLIANSVSLRWLPNLEIF 265
QY 285 -ELSQLNLSTNPLOKIQANOFDLYVKLSLGLGIEISNIOQRMFRPLMLSHIYFKFQ 343
Db 266 AELTQANLTY-PSHCCAFKPF-----KKNLEKNHLCNWS-----TIRQDEYFFEP-E 311
QY 344 YCGYAPHRSCKPNIDGISSLENLASSIQRVFVWVSAVTCFNGIFVICMRPIYSENK 403
Db 312 HCRDVIEV-SCYEPBDAPNCEDEDIMGFTFLRVLIWFISVLAIGNFVLLV--LLSHTK 368
QY 404 LYAMSTII--SLCCADCLMGITLFLVIGFDLPKFGVKNHQAOLMESTHCQLVSLAILST 461
Db 369 LTVPKFLMCHLAFADFCWGLIILIASVLDLTQSRYNBIEMQTGCGCTAGFLYFAS 428
QY 462 EVSVALTFLETKYICYIVPCFVCPGKCR---TIVLILYITGFIYVAFIPLSNKEFF 518
Db 429 ELSTVYLTATLERMHTITTYAMRLER--ECRLNHACCVMAGWVFSYLAALMPYIG---- 482
QY 519 KNYGTNGVCPPLHSEDTEISGAQIYVAIFLGINLAFTIYFVSYSMFTYSHQSAITA 578
Db 483 VSSYMKTSICLPM---DVEIVSSQVY-VMLLEFLNLVLAFCSCYVRITYIVRHPASVP 538
QY 579 TEIRNQVKKEMILAKRFFIYFDALCMPI-FVVKRFLSLQVEIPGTISWVVIPLPI 637
Db 539 D-----SADARAKKRAVAVLFTDPLCMAPISFAISALRQPLITYSHAKVILVLFPI 592
QY 638 NSALNPLIYLTTRPFKE 655
Db 593 NSCANPFLVAFIFKTFRQ 610

RESULT 15
Q9VYG0 PRELIMINARY; PRT; 334 AA.
AC 09VYG0;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, last sequence update)
DE 01-JUN-2001 (Tremblrel. 17, last annotation update)
DE CG4187 PROTEIN.
GN CG4187.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
```


